A

ACCESS TO ADVANCED ICT INFRASTRUCTURES

Testimony to the Senate's Subcommittee on Science, Technology and Space, Committee on Commerce, Science, and Transportation[1]

Neal Lane
Assistant to the President for Science and Technology, Director, Office of Science and Technology Policy
Washington, DC
March 1, 2000

Mr. Chairman and Members of the Subcommittee, thank you for this opportunity to testify about the important research and development investments proposed by S.2046, the Next Generation Internet (NGI) 2000 Act. These investments are a vital portion of the Administration's information technology (IT) research portfolio that strengthens and expands the important Federal networking research authorized, thanks to your sponsorship, by the NGI Act of 1998.

The Administration has been very encouraged by the active bipartisan support which both chambers of Congress have provided for efforts to strengthen our nation's investments in information technology research and development and we look forward to continued support for the exciting new work proposed in the administration's proposed FY2001 budget. Here in the Senate, your leadership, Mr. Chairman and that of the members of the Subcommittee, has been especially instrumental in helping your colleagues recognize that the advances in information technology which are so vital to the overall success of our nation's scientific and technical expertise, as well as to its economic prosperity, require a foundation of wise, sustained Federal research investments.

We are enjoying a time of unprecedented possibilities and prosperity, built on advances in science and technology enabled by Federal support for R&D. Creative businesses have translated the results of Federally funded advanced research into innovative products and services enjoyed today. This innovation has improved our quality of life, strengthened our national security, and unleashed an extraordinary era of postwar economic growth. Many of America's industries are now the most competitive and technologically advanced in the world. The Federal government has had an important role in sharpening our high-tech edge. Through policies such as investing in education, encouraging private-public partnerships, and limiting regulation of the Internet, the Administration has enhanced opportunities for scientific discovery and allowed innovation to flourish. Most importantly, as the President noted in his February 24 remarks to the Granoff Forum at the University of Pennsylvania, this Administration has worked to accelerate R&D at every level — pushing for an extension of the Research and Experimentation tax credit and increasing our national science and technology budget every single year over the last seven years.

The Nation Benefits from Federal IT R&D Investments

The case for sustained and adequate Federal investments in R&D is made most dramatically in the information technology sector. The President's Information Technology Advisory Committee (PITAC) notes that 'that the technical advances that led to today's information tools, such as electronic computers and the Internet, began with Federal Government support of research in partnership with industry and universities. These innovations depended on patient investment in fundamental and applied research.' The PITAC emphasizes, however, that continued Federal investment is essential to maintain this momentum. In their February 1999 report to the President, *Information Technology Research: Investing in Our Future*, the PITAC called for doubling Federal IT R&D investments over five years and expanding the existing coordinated interagency research programs to achieve a more balanced research portfolio. The Administration responded to the PITAC's proposals in FY 2000 with a major increase in IT research funding through the Information Technology for the Twenty-First Century initiative. We continue to build on the PITAC's recommendations with the programs recommended in the President's FY 2001 budget.

Although the dividends that our nation has reaped from past Federal investments in computing and communications research are well recorded, they are worth repeating. Federal support of IT R&D, leveraged by industry and academia, has led to technical advances which today are transforming our society and driving economic growth and the creation of new wealth. New computing, networking, and communications tools allow Americans to shop, do homework, and get health care advice online, and enable businesses of all sizes to join the international economy. Since 1995, more than a third of all U.S. economic growth has resulted from IT enterprises, and during the past decade, more than 40 percent of U.S. investment in new equipment has been in computing devices and information appliances. The IT sector is growing at double the rate of the overall economy and will soon account for 10% of the economy. Companies doing business on the Internet had an average market capitalization of $18 billion in 1999, more than 30 times the average market cap for all companies listed on the NASDAQ.

As computers, high-speed communication systems, and computer software become more powerful and more useful, IT penetrates deeper into our home, work, and education environments. Nearly half of all American households now use the Internet, with more than 700 new households being connected every hour. More than half of U.S. classrooms are connected to the Internet today, compared to less than three percent in 1993. In 1993, only a few technical organizations knew what an address like http://www.senate.gov meant, and today, there are nearly 13 million registered addresses. Today, more than 13 million Americans hold IT-related jobs, which are being added six times faster than the rate of overall job growth. Over 800,000 jobs were created by IT companies in the past year alone.

This astonishing progress has been built on a foundation of Federal agency investments in research conducted in universities, Federal research facilities, and partnerships with private firms. The Federal HPCC Program met its 1996 goals of demonstrating computers that perform a trillion operations per second and communication networks that transmit a billion bits per second. The Next Generation Internet initiative has exceeded its year 2000 goals by connecting more than 170 universities and other research centers at rates 100 times faster than those available when the project began and more than 15 institutions at rates 1,000 times faster. Such ultra-high-speed networks provide desktop-to-desktop connections nearly 20 million times faster than typical Internet connections to home computers.

The President's FY2001 IT R&D Budget

The President's FY 2001 budget reports all aspects of IT research — the base HPCC programs (including Next Generation Internet) and the new activities established by last year's Information Technology for the Twenty-First Century initiative — in a single integrated IT R&D program. The President is requesting $2.315 billion for IT R&D, $594 million more than last year's appropriations and a billion dollars more than the FY 1999 appropriation. The largest increases above FY 2000 funding are proposed for the National Science Foundation, which is leading the interagency effort (+$223M), the Department of Energy (+$150M), the Department of Defense (+$115M), the National Aeronautics and Space Administration (+$56M), and the Department of Health and Human Services (+$42M).

Agencies will continue to support the basic goals established in last year's initiative, focusing on fundamental research in software; development of information systems that ensure privacy and security of data and allow people to get information they want, when they want it, in forms that are easy to use; support for continued advances in high-speed computing and communications, including work needed to ensure that raw speed translates into usable speed; and work to understand the social, economic, and other impacts of IT with emphasis on ensuring that all Americans will benefit from these technologies. The U.S. research community responded to last year's call for research ideas with a flood of creative new proposals, a demand which far exceeded the supply of new funding in agencies such as NSF and DOD. As a result, with FY 2000 funding, NSF will start 25 small research centers and five larger centers.

TESTIMONY TO THE SENATE'S SUBCOMMITTEE ON SCIENCE, TECHNOLOGY AND SPACE, COMMITTEE ON COMMERCE, SCIENCE, AND TRANSPORTATION[1]

A crucial element is the continuing trend towards faster and cheaper computers.

As in previous years, the proposed IT research portfolio is based on coordinated, interagency investments which leverage expertise across agencies to give the best returns on those investments, both financial and technical. FY 2001 IT R&D priority areas include:

TEAMS TO EXPLOIT ADVANCES IN COMPUTING Expanded activities by NSF, DOE, NIH, NASA, and NOAA will support new partnerships where information scientists, mathematicians, and experts in areas such as medical research, weather modeling, and astronomy can work together to build tools for solving the Nation's most pressing information problems. These partnerships will advance information science and lead to research breakthroughs in application areas.

INFRASTRUCTURE FOR ADVANCED COMPUTATIONAL MODELING AND SIMULATION In FY 2001, NSF plans to establish a second terascale (five trillion operations per second) computing facility to support the civilian research community.

STORING, MANAGING, AND PRESERVING DATA Current networks and data storage systems are straining to support vast amounts of information. NASA's new earth observing satellite will generate data equivalent to three times the information in the Library of Congress every year. Research will include developing devices capable of storing a years output of such systems in devices the size of PC hard disks; searching data in a variety of formats including pictures, video, audio; and developing improved ways of filtering information, data mining, and tracking lineage and quality of information.

MANAGING AND ENSURING THE SECURITY AND PRIVACY OF INFORMATION Research will focus on systems that can ensure privacy and security without compromising speed and ease of use. DOE, for example, recently developed a prototype chip that can encrypt 6.7 billion bits per second. Work will accelerate in network protection and advanced encryption.

UBIQUITOUS COMPUTING AND WIRELESS NETWORKS This research will ensure that mobile and wireless systems can be integral parts of the Internet. These inventions will permit devices embedded in equipment, vehicles, portable or wearable devices such as medical monitoring equipment, and even kitchen appliances to identify themselves to networks automatically and operate with appropriate levels of privacy and security.

INTELLIGENT MACHINES AND NETWORKS OF ROBOTS Fundamental research in robots will help revolutionize our work and our lives – from earthmoving devices in hazardous environments to devices that fit inside blood vessels and help operating room surgeons to simple household robots. For example, NASA needs space probes that are smart, adaptable, curious, self-sufficient in unpredictable environments, and capable of operating in groups.

FUTURE GENERATIONS OF COMPUTERS New paradigms will use advances in quantum computation and molecular and nano-electronics to devise radically faster computers to solve problems previously described as 'uncomputable,' such as full-scale simulations of our biosphere or surgical simulations. Viewing cells as computational devices will help enable the design of next generation computers that feature self organization, self repair, and adaptive characteristics that we see in biological systems.

MORE RELIABLE SOFTWARE Software bugs and glitches continue to shut down airports, delay product shipment dates, and crash 911 emergency systems. Methods to design and test software need to be as productive and predictable as tools used to design and test aircraft and bridges.

BROADBAND OPTICAL NETWORKS DOD researchers have shown that optical networking can provide 1,000 times faster network backbone speeds. Improvements in optical switching and development of all-optical end-user access technologies will let users take full advantage of these speeds.

EDUCATE AND TRAIN A NEW GENERATION OF RESEARCHERS New investments will fund more researchers, who are critical to increasing both IT research and teaching, and support major research centers. Programs such as the teams to exploit advances in computing will provide opportunities to educate and train a new generation of researchers whose skills cross-disciplinary boundaries.

Large Scale Networking (LSN) R&D

The research priorities addressing network capabilities fall under the Large Scale Networking (LSN) R&D component of the coordinated, interagency IT R&D programs. Our ability to fully capture the future benefits of IT depends on learning how to build and use large, complex, highly-reliable and secure systems. The President's FY2001 budget proposes $334 million for LSN R&D, which includes:

- the LSN base programs in traditional networking research to support agency mission requirements
- the Next Generation Internet (NGI) initiative, and
- research in Scalable Information Infrastructure (SII)

LSN base programs explore long range fundamental networking research issues and transition developing LSN products into tools to support agency missions. Continuing the Federally-supported R&D responsible for the core technologies that made the Internet and Internet applications possible, LSN focuses on technologies needed by the Federal agencies, infrastructure to support agency networking, and networking applications development.

Since its inception in 1998, the Next Generation Internet (NGI) initiative has been a primary focus of

LSN, building on the LSN base programs to provide the networking research, testbeds, and applications needed to assure the scalability, reliability, and services required by the Internet over the next decade. The program has provided fast network testbed connections to 170 universities and other facilities, exceeding program goals for connecting 100 sites. It is now focused on two goals: providing revolutionary networking capable of operation a speeds a thousand times faster than typical systems operating when the program began, and providing key functionality for high speed networks including reliability, scalability, security, an ability to multicast, an ability to gracefully accommodate mobile wireless users and other users that may enter and leave the system, and other requirements of complex modern networks.

Scalable Information Infrastructure (SII) is the newest component of LSN. It was developed in response to PITAC recommendations for an expanded Federal role in networking R&D that includes interoperability and usability. The SII research goal is to develop tools and techniques that enable the Internet to grow (scale) while transparently supporting user demands. An integral part of LSN, SII R&D complements the LSN and NGI efforts. SII research will focus on deeply networked systems: anytime, anywhere connectivity; and network modeling and simulation.

Next Generation Internet 2000 Act

The Administration believes that the support for the LSN component of the coordinated, interagency IT R&D programs indicated in S.2046, the Next Generation Internet (NGI) 2000 Act is an important first step towards meeting our national needs for IT research. Fast, reliable, ubiquitous networks provide the lifeblood for a 21st century economy. They are essential for the conduct of business providing tools that can tie even the smallest businesses into international production and sales networks and let businesses of all sizes speed the rate they develop, test, produce, and market goods and services worldwide. Modern information networks are becoming essential elements of education and training, critical for providing safe air and highway transportation, and central for strategies aimed at boosting national productivity while minimizing the impact of economic activity on the natural environment. Fast, flexible, easily reconfigured networks are essential tools for our nation's military at peace, at war, and in the multiple peacekeeping and other tasks they are asked to provide. This is clearly a vital element of our national IT research portfolio, and the Administration welcomes the Subcommittee's support in gaining funding for this important research.

We feel strongly, however, that networking research must be conducted as an integral part of a program providing balanced investment in advanced software, high-end computing, high confidence systems, human-machine interface issues, and applications research which draw on innovations in both information science and research teams in areas such as advanced materials, climate and weather modeling, or astrophysics, as well as research into the social, legal, ethical and other issues raised by advances in information technology. This approach is consistent with the PITAC's directive to strengthen our Federal IT research programs by providing adequate funding for a complete and balanced IT research portfolio. We commend the Subcommittee for acknowledging in Section 3(1) of the bill the importance of supporting other IT research carried out by our Federal IT R&D programs. The language of the bill indicates, somewhat confusingly, that these activities should be authorized through the Next Generation Internet Program and the Large Scale Networking Program. However, the other elements of the Federal IT R&D program are complementary to, not subordinate to, the networking research authorized by the bill.

Networking research must be tied closely to research on the computers, the software, and the applications that drive them. Many of the most intractable problems in network research involve management of networks which may connect millions or even billions of nodes, providing high security and privacy at low cost in dollars or communication speed, and building systems which do not fail catastrophically when faced with component failures or hostile intrusion. All of these areas require close collaboration with researchers working software, the next generation of computers, and other parts of the information technology research program supported in our budget.

The President's FY2001 IT R&D budget presents all IT research, along with networking research, in a balanced R&D portfolio, as recommended by the PITAC. We hope that the Senate will support authorization for the entire range of information technology research as proposed by the President's budget and in accord with the PITAC's recommendations.

We were pleased to see the Committee's interest in providing the resources of information technologies to minority -serving institutions, rural communities and other underserved areas and groups. As you know, the Administration is seriously concerned about the nation's digital divide and its impact on the ability of these institutions to participate in our research enterprise. However, we believe that the bill is too prescriptive in providing resources for research on infrastructure for rural, minority and small colleges. Programs such as EPSCOR and the Minority Institutions Infrastructure already provide mechanisms through which these issues can be addressed. Also, starting with its new FY 2000 funding for IT R&D, the NSF has called on proposers to explore linkages with other institutions including HBCUs, Hispanic institutions, EPSCOR states and others to broaden the participation in the program. This strategy is used in many other ITR&D programs and links traditionally strong majority institutions with the strengths at HBCUs. We are concerned that specific set-asides provided through the legislation may not be the most efficient and productive way to provide greater opportunities for these institutions. We would like to work with the Committee to ensure that existing programs are strengthened to permit fuller greater participation in Federally-funded IT research and access to IT R&D resources.

We note that section 7 of the bill directs the National Academy of Sciences to conduct a digital divide study. The Administration believes this requirement should be deleted from the bill because it duplicates efforts already underway at the Department of Commerce. Commerce's National Telecommunications and Information Administration published the first 'digital divide' study in 1995. Its most recent study, 'Falling Through the Net: Defining the Digital Divide' (July 1999), has become the leading source of critical information on Internet access and computer usage. The NTIA study uses data collected by Commerce's Bureau of Census. The President's 2001 budget includes funding to permit NTIA to make this an annual study.

Many of the funding levels authorized by S.2046, as introduced on February 9, are consistent with those proposed for the LSN R&D programs in the President's FY2001 budget. One exception is that the proposed legislation does not appear to authorize funding for the National Oceanic and Atmospheric Administration (NOAA). NOAA is a long-time participant in the Federal LSN programs, including the Global Ocean Interactive Network (GOIN) demonstration project in March 1999 which linked U.S. ocean researchers with partners in Japan. Using links supplied by NASA, DoD, and NSF, NOAA's Pacific Marine Environmental Laboratory (PMEL) demonstrated the first NOAA applications over the NGI, including Ocean Share, a collaborative environment for oceanographic research, and 3-D tools using VRML to demonstrate the evolution of El Niño, fisheries larval drift, and fur seal feeding trips. Further research will include exploring methods of using advanced networks for aggregating the vast quantities of data from NOAA's satellite and radar weather sensors and multicasting the data to the nation's research community for the development of improved weather forecasting, developing tools to enhance collaboration among atmospheric scientists and oceanographers over the NGI, and increasing the robustness, security, and flexibility of networks for environmental research. We hope that the Subcommittee will modify its proposal to authorize funding for NOAA, as outlined in the President's budget.

Finally, although it received separate authorization in the NGI Act of 1998, the work on the Next Generation Internet initiative has always been an integral part of ongoing work in the Large Scale Networking component of the coordinated, interagency IT R&D program. This year, as noted above, LSN includes not only the base programs and NGI, but also expanded research in Scalable Information Infrastructure research. It appears that all of these elements, which are combined in the LSN R&D portion of the overall IT R&D program we plan to undertake, are authorized by S.2046. The Administration clearly prefers that the Committee take a more comprehensive approach to authorizing IT research. While the Committee takes this suggestion under advisement, we would urge you to refer to the programs authorized by the current proposed legislation as Large Scale Networking, rather than by the name of one of the program subcomponents (NGI).

I hope that we can work with the committee to make these modifications and resolve any other issues during the weeks ahead.

Conclusion

We thank the Subcommittee for its continued support of these vital research programs, first through the NGI Act of 1998 and now with the proposed NGI 2000 Act. These investments are an essential part of a larger, balanced portfolio of research developed according to the PITAC's directives for adequately funding our Federal IT research programs. The strong bipartisan support generated by these and complementary proposals allow us to invest in America's future and ensure its continued prosperity. We hope that we can work with the Committee to support the entire IT research portfolio proposed by the President. We believe strongly that this program provides a balanced program of research essential to the nation's prosperity and its ability to secure public benefits ranging from national security to environmental protection. I look forward to working with the Committee on these issues in the weeks ahead.

Notes

1 This text is online available at www.ccic.gov/legistation_testimony/it_march_lane.html.

Testimony to the Senate's Subcommittee on Science, Technology and Space, Committee on Commerce, Science, and Transportation[1]

Dr. Rita R. Colwell
Director, National Science Foundation, Arlington, VA
March 1, 2000

Introduction

Mr. Chairman, members of the subcommittee, thank you for allowing me the opportunity to testify on the National Science Foundation's role in fostering the next stages of the information revolution. I am pleased to be here today. This is a topic of utmost importance for the future of our nation's economy and the well-being of our fellow citizens. A healthy, long-term federal investment in high speed networking and information technology overall is critical if the United States is to remain a world leader – not only in science and engineering – but in our economy, national security, health care, education and overall quality of life.

My prepared remarks today will include a short history of NSF's support for cutting edge concepts in high-speed networking and their transfer to the private sector along with a brief discussion of the following topics:

- NSF's participation in the multi-disciplinary Federal Information Technology Research and Development Initiative (IT R&D) for which NSF is the lead agency;
- NSF's participation in the Next Generation Internet Program – an integral component of the IT R&D initiative – our cooperation with private industry through the rich transfer of new ideas to the private sector, our cooperation with the other NGI agencies;
- NSF's efforts to promote connectivity and access for all, including our efforts to improve connectivity for rural and minority-serving institutions and our strong support for cutting-edge education activities designed to ensure that our citizens will have the scientific, mathematical, engineering, and technological expertise needed to excel in tomorrow's knowledge-based economy.

NSF Support for High-Speed Networking: A Record of Accomplishment

Mr. Chairman, this subcommittee has long been a strong, bipartisan supporter of the federal investment in IT R&D. In the early 1980's, this subcommittee strongly encouraged NSF to invest in high-performance computing resources for the nation's academic scientists and engineers. The subcommittee also was a leader in the enactment of the High Performance Computing Act of 1991. This leadership continued with the passage of the bipartisan Next Generation Internet Act of 1998. With this backing from the subcommittee and the entire Congress, NSF has continued to support some of the most successful and innovative computer-communications concepts and technologies at their earliest, most experimental stages. NSF funded university-based supercomputer centers in the mid-1980's to provide academic scientists and engineers with access to state-of-the-art computing power. To facilitate access to the centers, NSF began a parallel effort in networking. It built on fundamental investments by DARPA in a more restricted environment, and resulted in the formation of the national NSFNET backbone network and regional networks connecting university students and faculty to the supercomputing centers. In a very brief period of time, NSFNET and the regional networks began performing important communication and information access functions in addition to supercomputer center access. Through this development and its subsequent privatization, the Internet industry was born.

Mr. Chairman, the story of NSF's longstanding support for backbone networks is now well known but it is only one example of how fundamental IT investments by NSF and other agencies have paid huge dividends for the nation. Support of fundamental networking research has received less publicity but is equally important to the future of information science and technology. For example, it was David Mills, an NSF grantee at the University of Delaware, who made it possible to have one Internet as opposed to a Tower of Babel of competing electronic networks. Mills developed the first widely-used Internet routers — the gateways and switches that guide the bits and bytes of data around the globe at the speed of light. That's why many people say NSF put the 'inter' in Internet. Today CISCO Systems – the premier maker of Internet router technology – now has a market capitalization of $454 billion dollars.

Knowledge Transfer
Not Just Technology Transfer

Innovations like the Internet router only occurred through sustained, long-term federal investments in information science and engineering by many agencies. One might think that these past successes assure us of an equally bright future. Unfortunately, in a fast paced, technologically-rooted information age, the worst thing we could do is rest on our laurels. The key point is that the IT R&D conducted by private industry – be it performed by large or small firms – is now primarily near-term and product-focused. There are many reasons for this trend. With increased global competition, increasingly rapid product cycling and high expectations from shareholders, IT industry managers tend to focus on activities that maximize short-term payoffs. Market pressures are often too great and technology changes too rapid to allow for major investments with a long-term perspective. When the subject of technology transfer is brought up, there is one aspect of the impact of basic research that is often overlooked – the role of NSF's investments in people. NSF's Engineering Directorate recently sponsored a set of studies on today's leading technologies: areas like cell phones, fiber optics, and computer assisted design. It's well known that the great majority of the seminal work in these areas was performed by private industry—at labs like Corning, AT&T, and Motorola.

Does that mean that NSF had no role? Hardly. When you go back and look at the work, a clear pattern emerges. Scientists and engineers who went to graduate school on NSF fellowships and research assistantships often brought the key insights to industry. In a number of cases, they became the entrepreneurs who created new firms and markets. To quote from the study – 'NSF emerges consistently as a major – often the major, source of support for education and training of the Ph.D. scientists and engineers who went on to make major contributions....' It is this transfer of people – the highly trained scientists and engineers supported by NSF and other agencies – that is making a tremendous impact on our knowledge-based economy. The NGI program is a tremendous success in this regard. In a preliminary review of the NGI program, the President's Information Technology Advisory Committee (PITAC) found that numerous NGI-funded scientists, engineers and students – first funded at universities – have gone on in just a few short years to found start-up companies with an estimated market capitalization of over $27 billion.

TESTIMONY TO THE SENATE'S SUBCOMMITTEE ON SCIENCE, TECHNOLOGY AND SPACE,
COMMITTEE ON COMMERCE, SCIENCE, AND TRANSPORTATION

Information Technology Research (ITR)

The impact of information technology on our society has been much wider and much more pervasive than anyone could have anticipated just a few years ago. Advances in computing, communications, and the collection, digitization and processing of information have altered the everyday lives of all our citizens. There is no question that as Internet growth has gone through the roof, IT has become the essential fuel for the nation's economic engine. Even the ever-cautious Fed Chairman Alan Greenspan has pointed to innovations in IT as the driving force behind our strong economic growth. The numbers speak for themselves. As Neal Lane has mentioned, more than a third of our economic growth in the past five years has resulted from Information Technology. IT investments have spurred an enormous upswing in worker productivity that has fueled the current economic boom. The challenge now is to sustain this record of success.

Last year, the PITAC concluded that federal support for long-term research on information technology has been 'dangerously inadequate.' In its words 'support in most critical areas has been flat or declining for nearly a decade, while the importance of IT to our economy has increased dramatically.' This has led to the government-wide initiative in Information Technology R&D for which NSF is the lead agency. The Information Technology Research Initiative at NSF will emphasize research and education on a broad range of topics. Focus areas include:

- Advancing computer system architecture; research on software, hardware, system architectures, operating systems, programming languages, communication networks, as well as systems that acquire, store, process, transmit, and display information.
- Improving information storage and retrieval; research on how we can best use the vast amount of information that has been digitized and stored.
- Connectivity and access for all; research that aims to overcome the digital divide separating the information 'haves' from the 'have-nots' and research on inequality of access to and use of computing and communications technology.
- Scalable Networks of Embedded Systems; As the scale of integration of systems that may be achieved continues to grow, systems must be designed with both hardware and software aspects treated from a unified point of view.
- Novel approaches; new models of computation and physical processes such as molecular, DNA and quantum computing. These efforts are deeply anchored in the mathematical and physical sciences and the biosciences.

Through our part of the multiagency IT R&D program, the Information Technology Research (ITR) initiative, NSF will seek to strengthen Education in IT, including:

- programs that provide scholarships, fellowships and traineeships;
- improved undergraduate research participation;
- encouragement of graduate students to participate in K-12 education and develop new curriculum; and
- research aimed at understanding the causes of underrepresentation of various segments of society in the workforce.

NSF will also increase research on Applications of IT across fields of science and engineering. This will also be a critical component of the ITR initiative. This includes simulation to tackle research problems across the frontiers of science and engineering. Important networking applications include:

- Collaboration Technologies
- Digital Libraries
- Distributed Computing
- Remote Operations and
- Security and Privacy issues.

Finally through the ITR Initiative, NSF will increase it's support for Infrastructure including the Next Generation Internet Program. Support for infrastructure will include:

- computing facilities ranging from single workstations to clusters of workstations to supercomputers of various sizes and capabilities;
- large databases and digital libraries, the broadband networking,data mining and database tools for accessing them;
- appropriate bandwidth connectivity to facilitate interactive communication and collaboration and software to enable easy and efficient utilization of networked resources; and
- networks of large and small physical devices.

NGI Connections at NSF: A Tremendous Success

Mr. Chairman, the NGI program has been a great success. Enabled by fundamental advances in optical networking under supported by DARPA and NSF, the number of very high performance networks has increased and the available bandwidth for research and education has had phenomenal growth. A diverse array of US universities in all 50 states now have high-speed connectivity thanks to NGI investments. In fact, many more institutions than originally anticipated now have high-speed access thanks to the program. Connectivity to Alaska and Hawaii has improved dramatically as well.

NSF's original goal under the NGI program was to connect 100 universities using the vBNS network and the Internet2 Coalition's Abilene network. Today NSF is excited that over 170 university connection awards have now been made. This includes over 40 universities in EPSCoR states – nearly one-quarter of the total. This increase in connectivity has resulted in interest in high performance networking in both academia and industry. It has had enormous impact on the knowledge transfer I mentioned earlier. Having so many more scientists, engineers and students from across the nation involved in high-speed networking activities has dramatically increased the available talent pool for industry. Universities form a rich, fertile proving ground for new network ideas and concepts that can be quickly transferred to the private sector. Without consistent federal funding, such a well-spring of ideas could run dry.

What's Next for NGI: The Next-Next Generation Internet

In marking our 50th anniversary, we are celebrating vision and foresight. The recently retired hockey-great, Wayne Gretzky, used to say, 'I skate to where the puck is going, not to where it's been.' Mr. Chairman, at NSF, we try to fund where the fields are going, not to where they've been. We have a strong record across all fields of science and engineering for choosing to fund insightful proposals and visionary investigators. It is our job to keep all fields of science and engineering focused on the furthest frontier. Our task is to recognize and nurture emerging fields, and to support the work of those with the most insightful reach. And, we prepare future generations of scientific talent.

Research communities will move to global unified archive systems.

In this tradition, NSF is looking at new directions for the NGI program. One trend is clear: high-speed fiber backbone networks are rich seed beds for new capabilities. Now that connectivity has been dramatically increased, new fundamental research problems must be tackled. In today's networked world, dramatic increases in backbone speed do not automatically translate into dramatic increases in performance. Many of these problems will not be easily solved without new, novel approaches.

Today, achieving high performance from end user to end user – the so called Broadband Last Mile Problem – remains difficult. Some commentators have remarked that the current situation is like having a four-lane highways beginning and ending with dirt roads. To increase backbone speed, efficiency and stability, we will need fundamental research into new middleware network service capabilities. This includes research in user authentication and verification, distributed computing services, and distributed storage services. Also, NSF will support research dealing with satellite and other wireless technology to help reach into areas where wireline and fiber are not possible or practical. We will also need research into new optical access technologies. In the future optical backbones will use more and more optical routing. Research is needed to discover how to appropriately extend the reach of these technologies. This will correspondingly extend the reach of networks and ensure that institutions not now taking advantage of high performance networking have the opportunity to do so.

Bridging the Digital Divide

This brings me to my last point. Today we find ourselves on a precipice – looking down into that worrisome gap known as the digital divide. We are all here today because we believe in the power of information technology to bring about the most democratic revolution in literacy and numeracy the world has ever known. We also know that if we're not careful, this same power could be economically divisive. We imagine universal connectedness, with talk of 'tetherless networks' that anyone could tap into anytime, anywhere. But we could also broaden the gap between the information rich and the information bereft. In our own nation, sociologists have identified groups whose access to telephones, computers, and the Internet lag far behind the national averages.

These information gaps appear among nations as well. Most of those who live in the Third World have never used a telephone. Our worldwide web is a thinly stretched one. Less than two percent of the world is actually on the web. If we subtract the United States and Canada, it's less than one percent. The report by the President's Information Technology Advisory Committee (PITAC) spells out some of these gaps. 'For instance,' says the committee, 'whites are more likely than African-Americans to have Internet access' at home or work. 'We expect there are similar gaps with other minority groups, such as Hispanics and Native Americans. Recent research ... suggests that the racial gap in Internet use is increasing.' In September 1999 NSF made a four-year $6 million award to EDUCAUSE to help minority-serving institutions develop campus infrastructure and national connections. The award addresses Hispanic, Native American, and Historically Black Colleges and Universities. The scope includes:

- Executive awareness, vision, and planning
- Remote technical support centers
- Local network planning
- Local consulting and training
- Satellite/wireless pilot projects
- New network technologies: Prototype installations
- Grid applications

Conclusion

To conclude Mr. Chairman, let me again thank you for holding this hearing so that we may exchange views on the future direction of this important area. Let me also restate NSF's willingness to work with you, the subcommittee and the full committee to ensure a robust federal IT investment including the NGI program. The PITAC report has raised important concerns over our lack of federal investment in fundamental IT research and we at NSF are responding to the challenge. We look forward to extending the federal IT partnership to help ensure U.S. world leadership in IT.

Thank you.

Notes

1 This testimony is online available at www.ccic.gov/legislation_testimony/it_march_colwell.html.

A European Research Area[1]

1 A Stock of material resources and facilities optimised at the European level

1.1 Networking of centres of excellence and creation of virtual centres World class centres of excellence exist in practically all areas and disciplines in Europe. Their exact specialities, however, are not always sufficiently well known outside the frontiers of the country in which they are established, especially by companies, which could usefully join forces with them. One of the criteria generally used to define the centres of excellence is their capacity to produce knowledge that can be used for industrial purposes.

Many problems of basic and applied research also need both a critical mass of financial and human resources and the combination of complementary expertise from specialists in other domains.

Mapping of European centres of excellence would make for a better transparency in this area. A very high level of performance could also be achieved by the networking of specialist centres throughout the countries of the Union. The forms of teleworking which electronic networks permit make it possible to create real 'virtual centres of excellence', in particular multidisciplinary and involving universities and companies.

To promote excellence, however, it is also necessary to ensure a sufficient level of competition between private and public research operators. Schemes to finance centres of excellence on the basis of competition have been put in place in several Member States. This formula could be applied to the European level, with collaboration between the Commission and the Member States.

1.2 Defining a European approach to research infrastructures Research infrastructures play a central role in the progress and application of knowledge in Europe. Radiation sources, computer centres and databases on molecular biology, for example, are operated increasingly by research teams from the public and private sectors. Facilities of this kind exist in all the Member States. Construction costs are high, often beyond the capacities of a single country, as are operating costs. And their potential is not always maximised.

Large-scale infrastructures have been constructed and are now being operated at European level. Furthermore, assessment of the need for new facilities is often made in a bilateral or multilateral framework. For its part, the European Union, has, for several years, been operating a programme of support for research infrastructures. So far measures within this framework have been restricted to providing support for transnational access to facilities, for the development of new instruments and equipment and for cooperation projects designed to improve the interoperability of installations and the complementarity of their activities.

What should be done now is to go a step further and develop a European approach to infrastructures, covering both the creation of new installations, the functioning of existing ones and access to them. An analysis of responsibilities (notably financial) on these three fronts should be made and plans to combine measures and means defined. Following on from the work carried out by the European Science Foundation (ESF) and the OECD, an accurate assessment should also be made of the needs to be covered at European level (including joint services).

A conference on the subject of research facilities in Europe will be organised in Strasbourg in the second half of the year 2000 by the Commission in conjunction with the European Science Foundation. This could provide the occasion for putting in place a framework in which to discuss these issues.

1.3 Better use of the potential offered by electronic networks

Electronic networks open up every new possibility of work to researchers: virtual laboratories; remote operation of instruments; quasi-unlimited access to complex databases. Created for the use of the scientific community, the Internet has also become the medium for multiple information and communication activities and has given rise to spectacular commercial developments. The World Wide Web, which was developed by a CERN researcher to cover the needs of physicists, is now used by tens of millions of people.

To meet the particular needs of research, which are constantly on the increase, specific networks are necessary. In the United States, broadband, high-speed facilities are now available to researchers, especially at universities. The recent Internet-2 and Next Generation Internet (NGI) initiatives launched in partnership by the scientific community, the public authorities and the private sector in the United States should increase these capacities even further.

To reduce the disparities in Europe in this area the Union is supporting an interconnection project of national telematics networks at progressively larger capacity levels: 34 Mbits/s, 155 Mbits/s now, and soon 622

Mbits/s, the ultimate objective being to achieve the order of magnitude of Gbits/s, at which some connections already operate in the United States.

To help Europe catch up quicker where electronic networks are concerned, the Commission proposed an e-Europe initiative at the Helsinki Summit, which sets ambitious objectives in terms in particular of interconnection at European level. It is accompanied by a timetable through to 2005. One of the aims is to promote maximum use of these networks by the community of researchers.

To increase the productivity of European research while helping to structure collaboration on a continental scale action will have to be taken in this context to encourage the use of electronic networks in the various fields of research in European as well as national research programmes: development of databases and access to advanced Internet services; promotion of the production of multimedia content and interactive uses; support for new forms of electronic collaboration of researchers ahead of the emergence of real 'virtual research institutes'.

At the same time it will be necessary to encourage researcher awareness-building and training campaigns at national and European levels on the possibilities created by information technologies and communications.

2 More coherent use of public instruments and resources

2.1 More co-ordinated implementation of national and European research programmes

Although they are often substantially funded, national research programmes are carried out largely independently of one another. This situation prevents the full benefit from being drawn from the material and human resources deployed. Research programmes in the Union exercise a certain co-ordinating effect on research activities in Europe. But this effect differs from one area to another. It is institutionalised in the case of fusion (which is covered by an integrated programme). It has an effect *de facto* in other areas, especially where there were still no structured programmes at national level when the action at European level was set in motion or in very specialist areas where there is not yet much expertise in Europe. The programmes of the Union should also have this impact more readily in areas where there is already appreciable integration of industrial efforts, like in aeronautics.

It would be right to go further in this direction by way of other mechanisms. The senior officials of the national research authorities in the Member States have decided to recommend the adoption of the principle of reciprocal opening-up of national programmes. The requisite measures will have to be taken to guarantee practical application. Mechanisms of reciprocal information and a global information system on the objectives and content of programmes plus the conditions for eligibility and participation should be put in place. This might also be extended to include applicant countries.

Convincing evaluation projects of national research activities by international panels made up largely of experts from the European countries have been completed in recent years in several countries, in Portugal and Germany, for example. Initiatives of this kind have to be encouraged.

In this area the Commission can play the role of initiator and catalyst by providing the Member States with the logistical means and legal instruments best suited to co-ordinating research activities undertaken in Europe.

2.2 Closer relations between scientific and technological cooperation organisations in Europe

Over the last twenty to thirty years or so, alongside the European research programmes (and even before them), a series of organisations for European scientific and technological co-operation have been created in an intergovernmental framework) (ESF, ESA, EMBO, EMBL, CERN, ESO, ESRF, ILL, EUREKA, COST).[2]

Co-operation has developed between them and with the research programmes of the Union essentially on a bilateral basis (co-operation between the Union and EUREKA, ESA and ESF in particular).

These organisations play an important role on Europe's science and technology scene. They are now facing common problems (financing, integration of researchers from central and eastern European countries, dialogue with the United States). It would be useful to provide them with a framework in which they could discuss their respective roles on the European scientific and technological scene and their relations between one another and with the Union.

Priority must be given to establishing the conditions for political consultation between these organisations. This could be achieved by way of a council of their senior officials meeting at regular intervals. This would also give Europeans and outside observers a more coherent image of the Europe of science and technology.

3 More dynamic private investment

3.1 Better use of instruments of indirect aid to research Increasing use is being made in the world of instruments for indirect support, especially fiscal measures, in order to stimulate private investment in research and development and to create researcher and technician posts in companies. In the United States and Canada, interesting long-term support schemes for start-up companies are in place, for example. In Europe the mechanisms used in the various countries are very diverse. Some Member States make sustained use of them, others far less.

User-friendly information systems need to be developed on existing mechanisms. The exchange and spread of good practices should also be encouraged in order to stimulate private investment in research, particularly among SME's, and innovation.

The different situations between countries and regions in this area can affect competition between them in many ways and create conditions that are more, or less, conducive to investment in research and innovation. Where the measures employed have an element of State aid about them, Community rules on State aid should always be respected.

3.2 Development of effective tools to protect intellectual property The current European system of patents, as operated around the European Patents Office and the national offices, is based on the issue of national patents which are valid only in the Member States in which they are issued. This system is costly and the high cost of patents is broadly believed to be one of the major obstacles to widespread use of patents in Europe. The Commission therefore plans to propose the creation of a standard Community patent to cover all of the European territory. At international level the Commission will endeavour to adapt the TRIPS agreements on intellectual property to new technological developments.

Stunning advances in information, and computation are transforming our world

It is important for research in Europe for the European patent to be started up as soon as possible. It must be readily affordable and comparable in cost to a European patent covering a limited number of countries. Efforts need to be made in particular to reduce the costs of translation. The Commission is also keeping a close eye on work carried out by the European Patent Organisation as part of the revision of the Munich Convention in order to see how the effects of disclosures prior to filing can be taken into account by European patent law.

To increase the impact of research efforts undertaken in Europe in terms of innovation, the relevance and consistency of the intellectual property arrangements used to implement public research programmes should also be improved.

The protection of intellectual property can be achieved by other means than patents. In addition to the initiatives taken in the First Action Plan for Innovation in Europe, information systems and systems for exchange of good practices in this field could be put in place by national and European support organisations for research and innovation.

3.3 Encouragement of the creation of companies and risk capital investment

The creation of high tech companies by researchers, or with researchers having a stake in the capital, is still fairly low-key in Europe. Measures taken in recent years at regional level, such as the creation of technology parks and business incubators, or by certain Member States, such as changes in the status of public sector researchers, have had a positive effect in this respect. These could be completed by other initiatives.

Europe also suffers notoriously from too low a level of risk capital investment in high tech sectors. Positive changes have been observed for some time now. Some 650 companies are now quoted on the new European markets (Euro-NM, EASDAQ and AIM). That said, this is eight times fewer than in the United States. In extending the first action plan for innovation in Europe, the Commission has, in recent years, taken a series of initiatives in this area, several of which (e.g. I-TEC project) are being implemented in conjunction with the European Investment Bank (EIB). In 1999 it presented two communications on this subject.[3] As part of the e-Europe initiative the Commission recently proposed a plan of action designed to establish an inventory by March 2000 of existing instruments at Union level.

computing,

rapidly.

Several national research centres and the JRC have joined forces to provide innovative start-up companies with the technical support and expertise they need to develop. Initiatives of this kind need to be stepped up.

Initiatives should also be encouraged to bring scientists, industrialists and financiers at all levels into contact. This could be achieved in conjunction with the national and European research programmes, preferably on a combined basis. Promising experiments have been completed along these lines, like the "Investment Forum" in the field of information technology and communications and the creation of the "Biotechnology and Finance Forum".

4 A Common System of Scientific and Technical Reference for Policy Implementation

4.1 Developing the research needed for political decisions

Science and technology play an increasingly important role in the implementation of public policies, particularly Union policies. They are involved in various forms in the drafting of regulations and can be found more and more in the policy-making process, at the heart of trade negotiations and at the centre of international discussions in fields such as, for example, safety in its various forms or the various aspects of sustainable development.

The European research system must be organised in such a way as to preempt and take account of needs arising at the different stages of implementation of public polices: drafting, decision-taking, implementation, monitoring. Policy-makers must be able to draw on precise knowledge which is as complete as possible and constantly updated and validated.

Accordingly, research directly undertaken by the Commission must tie in with the major concerns of the individual and the decision-makers, such as environmental protection, food safety and chemical products or nuclear safety.

The results of research undertaken as part of European programmes should be systematically exploited in support of the various Union policies and all the Union's research activities better co-ordinated in this respect.

A reliable and recognised system of validating knowledge and methods of analysis, control and certification also needs to be put in place and centres of excellence in Europe in the fields concerned networked.

4.2 Establishment of a common system of scientific and technical reference

When drawing up regulations or when faced with emergency situations, policy-makers, especially at European level, are confronted with complex problems where the stakes are high. Citizens and economic and social operators must be guaranteed greater safety whilst resolving conflicts between categories of actions with often divergent nterests. As the Commission underlined in the White paper on food safety,[4] the Union must re-establish the confidence of the public and consumers in food (they way it is produced, regulated and controlled).

In Europe the way expertise is provided for decision-makers differs according to country and subject matter. Authorities established at European and national levels abound. Experts are also forced to leave the ground of solely scientific consideration. The way they assess the problems and their recommendations bear the imprint of their iscipline, their areas of activity or the community to which they belong.

By aligning methods, harmonising procedures and comparing results, a common system of reference needs to be established at Union level. Given its institutional proximity to the development of the Union's policies and its independence of national and private interests, the JCR could, in line with its mission, play a significant role in the development of a European scientific and technical reference area. This would be built up on the basis of national reference centres, European agencies, the various scientific committees and the organisations established at European level, such as the Food Safety Authority, free of industrial and political interests and open to public enquiry and scientifically recognised, which the Commission has suggested be established by 2002 following broad consultation.

Notes

1 Excerpts from "Towards a European Research Area", COM (2000) 6, Brussels, 18 January 2000.

2 ESF: European Science Foundation; ESA: European Space Agency; EMBO: European Molecular Biology Organisa-

tion; EMBL: European Molecular Biology Laboratory; CERN: European Organisation for Nuclear Research; ESO: European Southern Observatory; ESRF: European Synchrotron Radiation Facility; ILL: Institut Laue-Langevin; COST: European Cooperation in the field of Scientific and Technical Research.

3 COM (99)232 and COM (99)493.

4 COM (99)719.

Information Technology Research: Transforming Our Future

Dr. Ruzena Bajcsy
Assistant Director, Computer and Information Science and Engineering Directorate
National Science Foundation

The Technology Revolution. The Information Age. The New Economy. We are living in exciting times today, and no one is quite sure how to describe them accurately. What is certain however is that stunning advances in computing, information, and computation technologies, particularly over the last two decades, are transforming our world rapidly in ways that are often difficult to comprehend. Like the discovery of fire, the invention of movable type, and the harnessing of electricity, these transformations will profoundly influence how we live our lives in the future – but their effects can be maddeningly unpredictable.

In the U.S., many of these transforming technologies, including the Internet itself, grew out of basic research supported by U.S. Government agencies such as the Defence Advanced Research Projects Agency (DARPA) and the National Science Foundation (NSF). But in recent years, Federal investments in crucial high performance computing and communications research has not kept up with rapid developments in these technologies. To address this problem, in February 1997 the Clinton Administration established a bipartisan Presidential Information Technology Advisory Committee (PITAC), comprised of technology leaders from business and academia, to advise the U.S. Government how to bolster its sagging investments in critical Information Technology Research and Development (IT R&D).

In its path-breaking 1999 report to the President, *Information Technology Research: Investing in Our Future,* the PITAC addressed the ways in which technology is transforming modern society and issued a series of forward-looking recommendations designed to support the research community and ensure that the benefits of technological education and innovation would be made available to all citizens. In addition, it underlined the continued importance of Government investments in the research and development of leading-edge technologies that might take decades to achieve fruition—precisely the kind of long-term R&D that increasing numbers of private sector businesses, driven by the incessant demands of the competitive global economy, no longer find it economically feasible to support.

The PITAC outlined ten dramatic ways in which information technology will transform or is already transforming our society. These include:

Transforming the Way We Communicate

The Internet is now the heart of the communications revolution, but its performance falls considerably short of anytime-anywhere instant communication. The present Internet must be expanded in scale to anticipate growth. Communications technologies must be simplified and networks made more robust. The evolution of global networking poses international problems as well—problems that extend beyond technology into the realm of international law. How do we maintain international boundaries, customs, and traditions? How do we protect international intellectual property, copyrights, and patents from theft or unauthorised use? How do governments and individuals ensure the security of information and transmission? These are all questions that need answers before we can full achieve and take advantage of the Internet, which is still in a state of relative infancy.

Transforming the Way We Deal with Information

Humans can now interface with information in far more ways than was formerly the case. In addition to television and the venerable printed word, users can now access information online in a variety of ways. But these interfaces are still primitive and clumsy. There are easier ways to develop information besides employing a keyboard. Search technologies are also inefficient and clumsy. Searches currently produce a bewildering amount of information that is difficult to sort. How can we improve the interface with information? The answer undoubtedly lies with the development of new multi-modal human computer interaction technologies such as speech, touch, and gesture recognition and synthesis. With new ways of accessing information, we can achieve equal access by novices and experts, regardless of physical condition or global location.

Systems also require improvements in data access methods. Research issues include network reliability and bandwidth, scaleable software support, database structure and retrieval algorithms, robust and secure access, as well as quality of audio and video.

Interfaces are still primitive and clumsy.

Transforming the Way We Learn

Information Age technologies are serving to improve the information infrastructure but we need to further improve underlying software technologies to enable fast, easy development of new educational materials and support dynamic modification and maintenance of these materials. We need also to determine which educational needs can be fulfilled by computing and communications technologies, and which needs can be fulfilled by traditional methods. Lifelong learning will help citizens learn and use new technologies effectively in personal and professional lives.

Transforming the Practice of Health Care

Medical care is an area that has vast potential for technological improvements. U.S. Government agencies such as the National Institutes of Health (NIH) and the Agency for Healthcare Quality and Research (AHRQ) are already developing a national infrastructure for electronic medical records and health system intranets for data-sharing. However, there remain critical problems of privacy and security of patient records as well as the difficult technologies of data sharing. Further research is also needed in making telemedicine work for both patients and physicians, and new research is needed in remote visualisation and robotics—technologies that can take medical expertise, readily available in large urban areas, to isolated and under-served rural populations that do not currently have access to such services. In this area as well, high-reliability, low-latency communications are needed to support health care applications such as telepresence surgery.

Transforming the Nature of Commerce

The Internet has already changed how business is transacted worldwide. It is clear that new technologies can get companies closer to customers and reduce paperwork, purchasing costs and delivery time. Again, privacy and security of corporate and individual information are critical topics, as are the thorny problems of international trade relations in the Information Age where proprietary information that was once difficult to access can become common knowledge in seconds.

Transforming the Nature of Work

As many as 15 million U.S. workers will become telecommuters over the next decade. This will enhance productivity, and provide organisational flexibility and environmental benefits. However, to achieve these desirable results, truly high speed networking capacity will be needed, collaborative software technologies will be critical and the social and economic implications must be studied so that the worker skill base can be rapidly and effectively updated and renewed.

Transforming How We Design and Build Things

New technologies are already increasing productivity, reducing the cost of goods sold, improving the quality of merchandise, maintaining maximum planning and manufacturing flexibility, and reducing design cycle time. Improved high end technologies are still needed for concept design, simulation, analysis and data mining and the National Institute of Standards and Technology (NIST) is currently supporting important work in these areas. There is also a need for improvements in planning and scheduling, purchasing, investment, and cost analysis. An exciting development in this area is the growing probability that networked computers can allow simultaneous, interactive modification of standard products to meet specific customer needs.

Transforming How We Conduct Research

Research problems are becoming more complex and interdisciplinary in nature. But keeping pace with these problems, high-speed computers and networks are enabling discoveries across a broad spectrum. Innovative methods for collaboration around the globe are being created and promoted. Key technologies in this area are high-end computing for modelling complex physical phenomena, advances in collaborative environments, visualisation of complex datasets, innovative data-mining techniques and improved management of very large datasets and databases

Transforming Our Understanding of the Environment

We need to conduct further research into climate modelling to improve accuracy of local and regional forecasting, disaster management and support for national and international energy and environmental policies. Progress in this area depends on improved computational methods requiring order of magnitude increases in computing capability to deal with immense problems of time and space. The National Aeronautics and Space Administration (NASA), the National Oceanic and Atmospheric Administration (NOAA) and the Environmental Protection Agency (EPA) all conduct important IT R&D in these areas, and these efforts must be continued and increased.

Transforming Government

We need to employ new technologies in all government institutions to make them more efficient and responsive to constituents. Challenges in this area include improvements in systems and data access methods, high performance data storage and tools to locate and present information. Again, robust, reliable and secure networks are critical. Improvements must be available to *all* citizens and barriers to access must be surmounted.

To accomplish PITAC's far-reaching recommendations, the U.S. Government will need to emphasise new priorities in its high-end R&D investments including; support for fundamental research in software R&D methods and component technologies, fundamental research in human-computer interfaces and interactions and fundamental research into information capture, management, analysis, and availability. In fact, software research will need to be a substantive component of every major IT research initiative.

U.S. Government agencies are being challenged to develop new management structures to ensure that important national goals for information technology R&D are addressed. This management structure must connect complementary efforts across research disciplines as well as funding agencies. Most importantly, PITAC has called upon the NSF to assume a larger leadership role in basic IT R&D activities. NSF and other agencies are being asked to fund research into innovative computing technologies and architectures and software to improve the performance of high-end computing.

Government-supported high-end computing research should strive to attain a sustained petaops/petaflops on real applications by 2010 through a balanced approach to hardware and software technologies. An additional priority for U.S. Government agencies is to work to ensure that the average citizen is not left behind in this era of rapid technological advancements. All citizens must have equal access to information technology, and IT education and training must be augmented and expanded if this goal is to be accomplished.

A dramatically expanded IT research agenda is essential for continued and sustained economic growth as well as national and international security. PITAC's strong and detailed recommendations offer a good start on a new, augmented program of 21st century IT R&D innovations, but there is much work remaining to be done and many important discoveries yet to be made, both in the U.S. and in the international science community. As an exciting and promising new century begins to unfold, the only certainty is that the Future is Now.

Lessons from the Past, Global Challenges Ahead

Karel Vietsch
TERENA

Four years ago I had the pleasure at the first OECD conference on 'The Global Research Village' of speaking about research networking. I talked about the status of global research networking at that time and I pointed out the challenges ahead. In particular I addressed the questions 'Who shall pay for the research networks?' and 'What is the role of governments?'. Now I would like to revisit the status of research networking worldwide. There have been enormous changes in the past four years, mostly for the better. There are however, new and important challenges that need the attention of research networking organisations, funding bodies, research and higher education institutions, and governments.

Nowadays research and education depend increasingly on electronic media and networks. In technologically and economically developed countries, networking services for research institutes and educational establishments are provided by research and education networks. These networking organisations collaborate at an international level, thus creating a high-quality information and telecommunications infrastructure. TERENA is the association in which the research and education networking organisations from countries in and around Europe collaborate. TERENA's work falls into four main categories:

- representing the interests and opinions of its member organisations;
- developing, testing and promoting new technologies and services through the TERENA Technical Programme ;
- organising conferences, workshops and seminars;
- acting as a cradle for new initiatives and services.

At the time of the first 'Global Research Village' conference, the costs of research networking were the biggest problem. The prices charged by public network operators were very high and not related to real cost. Prices also varied widely from country to country, and in Europe prices were typically 10 times higher than in North America. The liberalisation of the telecommunications markets in the European Union member states that started in 1998 has had an enormous impact. Prices have been going down by a factor of 4 every year. For international connections in Europe the average price per Megabit-per-second per year is only some 4% of what it was 6 years ago.

At the same time national governments, funding bodies and universities have come to understand the importance of research networking. In the past few years a number of big initiatives have been launched, like Internet2 in the United States, CANARIE's all-optical network in Canada and the Gigaport project in the Netherlands. This has given a big stimulus to the development of the research networking infrastructure, services and applications. The typical capacity of research networks now is 50 to 500 times higher than four years ago. This Gigabit networking involves new technologies and opens up opportunities for new applications such as those enabled by streaming media or by computational and data grids.

There are still a great number of challenges ahead. Users will only be able to benefit fully from these new networks with very high capacities, if new and advanced applications and services are developed. Much more work needs to be done in these areas. With very high capacity infrastructures in place, content also needs to be in the focus of attention. More capacity will only lead to information overflow if no better ways are developed to find and access information. Finally it should be noted that countries outside the OECD have not been able to keep up with developments in the most advanced countries over the past years. The technology gap has only become wider, and advanced research networking facilities in the countries of the former Soviet Union, in South Asia, in Africa and in large parts of Latin America seem further away than ever.

New Frontiers for Research Networks in the 21st Century

Robert J. Aiken
Manager, University Research, Office of the CTO, Cisco Systems, Inc.

DISCLAIMER The ideas, comments, and projections proffered in this paper are the sole opinions of the author, and in no way represent or reflect official or unofficial positions or opinions on the part of Cisco Systems, Inc. This paper is based on my experience designing and managing operational international research networks, as well as being a program manager for network research, during the formative years of the Internet (i.e., my tenure as a program manager for the United States Government's National Science Foundation and the Department of Energy), and my recent experience within Cisco working with next generation internet projects and managing its University Research Program. Many of the examples that I cite in this work are based on the development and deployment of the US based Internet and research networks; although, the lessons learned in the US may also be illuminating elsewhere.

GRATITUDE I would like to thank my friend and colleague, Dr. Stephen Wolff, of the Office of the CTO, Cisco Systems Inc., for many good suggestions with respect to improving the content and presentation of this paper; but, mostly for his good humored authentication on my history and facts.

Abstract

A famous philosopher, Yogi Berra[1], once said 'Prediction is hard. Especially the future.'. In spite of this sage advice, we will still make an attempt at identifying the frontiers for research networks. By first examining and then extrapolating from the evolution and history of past research networks we may be able to get an idea about what frontiers face research networks in the future. One of the initial roles of the research network was to act as a testbed for network research on basic network protocols, mostly focusing on the network layers one through four (i.e. the physical, data link, network , transport, and network management layers), but also including basic applications such as file transport and e-mail. During the early phases of the Internet the commercial sector at this time could not provide the network infrastructure sought by the research and education communities. Consequently research networks evolved and provided backbone

and regional network infrastructures that provided production quality access to important research and education resources such as supercomputer centres and collaboratories[2]. Recent developments show that the majority of research networks have moved away from being test-beds for network research and have evolved into production networks serving their research and education communities. Its time to make the next real evolutionary step with respect to research networks and that is to shift our research focus towards maximising the most critical of resources, people.

Given the growth and maturity of commercial service providers today, there may no longer be a pressing technical need for governments to continue to support pan national backbone networks, or possibly even production like national infrastructures, for Internet savvy countries. Since commercially available Virtual Private Networks (VPNs) can now easily support many of the networked communities that previously required dedicated research networks, government and other supporting organisations can now support their research and education communities by providing the funding for backbone network services much as it does for telephony, office space, and computing capabilities, i.e., as part of their research award. However, there may be valid social, political, and long-term economical reasons for continuing the support for such networks. For instance, a nation may decide that in order to ensure its economic survival in the future it wishes to accelerate the deployment and use of Internet technologies among its people and thus they may decide to subsidise national research networks. In addition, it should be noted that VPNs often recreate the 'walled' separation of communities, which previously was accomplished through the hard muxing of circuits. But, in order to make technical advances in the e-economy, Governments should now focus on supporting the evolution of intelligent and adaptable edge and access networks. These, in turn, will support the ubiquitous computing and persistent presence environments that will soon be an integral part of our future Internet based economies.

The USA's recently expanded National Science Foundation (NSF)[3] research budget and DARPA's[4] prior support of middleware research are good examples of moving in the right direction. The Netherland's Gigaport[5] project, which incorporates network and application research as well as an advanced technology access and backbone network infrastructure, is a good example of how visionary research networks are evolving.

Just as Internet technologies and network research have matured and evolved, so should the policies concerning the support of research networks. Policies need to be developed to again encourage basic network research and the development of new technologies. In addition, research networks need to encourage and accentuate new network capabilities in edge networks, on campus infrastructure, and in the end systems to support the humans in these new environments. This paper will mainly focus on the future of research networks in e-developed nations; but, this is not to diminish the need or importance for e-developed nations to help encourage the same development in network challenged nations.

Context and Definitions

Before delving into our discussion we first need to define a few terms. These definitions will not only aid in our discussion, but may also help to highlight the role and function of various types of research networks. The most important terms to define are those of 'network research' and a 'research network', both of which often get interchanged during discussions concerning policy, funding, and technology.

For the purposes of this paper we use the term network research to mean long-term basic research on network protocols and technologies. There are many types of network research that can be roughly categorised into 3 classes. The first category covers research on network transport infrastructure and generally includes research on the OSI model layers 1 through 4 (i.e., the physical, data link, network, and transport layers) as well as research issues relating to the interconnection and peering of these layers and protocols. We will refer to this class of research as transport services. The second class is constituted of research covering what can nominally be referred to as middleware[6]. Middleware roughly includes many of the services that

The majority of the
of the network
will live at the edges.

were originally identified as network layers 4 through 6. Layer 4 is included because of the need for interfaces to the network layer. In addition, it nominally includes some components, such as email gateways or directory services, which are normally thought of as being network applications, but which have sub-components that may also be lumped into middleware. Given that the definition of middleware is far from an exact science, we shall say that middleware depends on the existence of the network transport services and supports applications. The third area covers research on the real applications (e.g., e-commerce, education, health care, etc.), network interfaces , network applications (e.g., e-mail, web, file transfer, etc.), and the use of networks and middleware in a distributed heterogeneous environment. Applications depend on both the middleware and transport layers. Advanced applications include Electronic Persistence presence (EPP) and Ubiquitous Computing. EPP, or e-presence, describes a state of a person or application as always being 'on the network' in some form or another. The concept of session based network access will no longer apply. EPP assumes that support for ubiquitous computing and both mobile and nomadic network exists. Ubiquitous computing refers to the pervasive presence of computing and networking capabilities throughout all of our environments, i.e., in automobiles, homes, and even on our bodies.

A research network, on the other hand is a production network, i.e. one aspiring to the goal of 99.99999% 'up time' at layers 1 through 3, and which supports various types of domain specific application research. This application research is most often used to support the sciences and education but can also be used in support of other areas of academic and economic endeavour. These networks are often referred to as research networks (RNs) or Research and Education (R&E) networks. To aid in the discussion of this paper we will further classify these RNs based on their general customer base. Institutional Research Networks (IRNs) are networks that support universities, institutes, libraries, data warehouses, and other 'campus' like networks. National Research Networks (NRNs)[7], such as the Netherland's Gigaport or Germany's DFN networks, support IRNs or affinity based networks. Pan National Research Networks (PNRNs) interconnect and support NRNs. An example of a couple of current production PNRNs are Dante'sTen-155 and the NORDUNET[8] networks. For the purpose of this paper we will also classify the older NSFNETs, vBNS, CANARIE's CA*NET 3[9], and the Internet 2[10] Abilene networks as PNRNs because in terms of scale and policy they address the same issues of interconnecting a heterogeneous set of regionally autonomous networks (e.g. NSFNET's regionals and I2's gigapops) as do the PNRNs.

There also exists a hybrid state of RN. When we introduce one or more advanced technologies into a production system, we basically inject some amount of chaos into the system. The interplay between the new technologies and other existing technologies at various levels of the infrastructure, as well as scaling issues, can result in unanticipated results. Research quality systems engineering and design is then required to address these anomalies.

For the purposes of this paper we define a few more terms we use throughout this paper. The term Virtual Private Networks (VPN) is used in the classical sense, i.e. a network tunnelled within another network (e.g. IP within IP, ATM VCs, etc.), and it is not necessarily a security based network VPN. Acceptable use policy (AUP) refers to the definition of what type of traffic or use is allowed on a network infrastructure. Conditions of Use (COU) is basically another version of AUP.

Introduction

During the early phases of the evolution of research networks and the internet, national research networks were building and managing backbone networks because there was a technical reason to do so. Governments supported these activities, because at the time the commercial sector could not do it and the expertise to do so resided within the R&E community. Much of the research or testing of this time was still focused on backbone technologies as well as aggregation networks and architectures. Research networks started out by supporting longer term risky network research and quickly evolved to support shorter term no-risk production infrastructure. The research during the ARPANET and early NSFNET phases of the Internet focused on basic infrastructure protocols and technologies. These services are now commodity services and both easily and cost effectively available from the commercial sector. We have come a long way since then. Except for a few universities and research centres, the commercial sector now dominates R&D in the backbone technology space. Commercially provided VPNs can now cost effectively support the majority of the requirements of the R&E communities. Given the current domination of R&D in backbone technologies by the commercial sectors, as well as the need to address true end to end services, it is time that network research and research networks realign their focus onto the research and development of end system and campus and edge network technologies. The majority of the intelligence of the network (e.g. quality of service, security, content distribution and routing, etc.) will live at the edges, and in some way will be oblivious to the backbone service over which it will operate. In addition, in order for applications to be able to make use of this network, intelligent RNs need to be able to provide the middleware and services that exist between the application and the transport systems. The real future for most RNs is in helping to analyse and identify, not necessarily run and manage, advanced network infrastructures for their R&E communities.

One of the problems faced by the R&E community is how to obtain support from their governments and other supportive organisations (both for-profit and non-profit). In attempts to support advanced applications and end user research, organisations and governments may be convinced into supporting RNs, which end up providing commodity services and competing with the commercial sector. One reason that this can occur is that governments often wish to see results very quickly in order to justify their support of the research community; but, by doing so they drive the recipient researchers and research network providers to focus on short term results and abandon basic long term research. This pressure from the supporting organisations can also force researchers to compete in a space, i.e. transport layers, for which industry may be better suited and adapted in both scale and time. Another issue facing today's research networks is that many of the R&E community, who once would endure down time and assume some risk in trade for being part of an experimental network, are now demanding full production quality services from those same networks. Subsequently, the RNs are then being precluded from aggressively pursuing and using really advanced technologies that may pose a risk. And finally, many times research networks, science communities and researchers claim they are doing network research, when in reality they are not, because they wish to have decent network connectivity, and they assume that this is the only way to get funding and support for good network connectivity with which to support their real research objectives. All of these issues have driven RNs at all levels into difficult positions. RNs need to be able to again take risks if they are to push the envelope in adopting new technology. Likewise, it is also valid to provide production quality network transport services to support research for middleware, network application (e.g. collaborative technologies), and R&E application (e.g. medical, sciences, education, etc.) research. All of these requirements need to be addressed in the manner most expedient and cost effective to the government or organisation providing the support.

All research carries with it a certain amount of risk. There is theoretical and experimental research. Some research is subject to validation; some is *retrospective* – e.g., examining packet traces to verify the existence of non-linear synchronisation, but some is *prospective* and involves reprogramming network resources and any reprogramming is susceptible to bugs. The amount of risk often depends on the area of research undertaken. The lower down in the network structure that one performs experimental research, the harder it is to support this research and still maintain a production like environment for the other researchers and applications; yet we need to provide support for all levels of experimental research, as described in MORPHNET[11]. The ideal environment would support applications that could easily migrate from a production network to one prototyping recent network research, and then back again if the experiment fails. Recent advances in optical networking show promise in realising this goal, but there are many technical and policy based challenges yet to be addressed.

ARPANET and Early NSFNET Phase: 1980s

The ARPANET, one of the many predecessors of today's Internet, was a research project run by researchers as a sandbox where they could develop and test many of the protocols that are now integral components of the Internet. Since this was a research network that supported network research, there were times the network would 'go down' and become unavailable. Although that was certainly not the goal, it was a reality when performing experimental network research. This was acceptable to all involved and allowed for the quick 'research to production' cycle, now associated with the Internet, to develop. The management of the network with respect to policy was handled by the Internet Advisory Board (IAB), which has since been renamed the Internet Activities Board, and revolved around the actual use of the network as a research vehicle. The research mainly focused on layers 1-4 and application research was secondary and used to demonstrate the underlying technologies.

At the end of the 1980s, the Internet and its associated set of protocols, rapidly gained speed in deployment and use among the research community. This started the major shift away from research networks supporting experimental network protocols towards RNs supporting applications via production research networks, e.g. the mission agencies' (i.e. those agencies whose mission was fairly well focused in a few scientific areas) networks at DOE (ESnet) and NASA (NSInet). At the same time the NSFNET was still somewhat experimental with the introduction and use of 'home grown' routers, as well as with pioneering research on peering and aggregation issues associated with the hierarchical NSFNET backbone. It also focused on issues relating to the interconnection of the major agency networks and international networks at the Federal Internet Exchanges (FIXes), as well as the policy landscape of interconnecting commercial email (MCIMail) with the Internet. The primary policy justification for supporting these networks (e.g. ESnet, NSInet, NSFNET) in the late 1980s was to provide access to scarce resources, such as supercomputer centres; although the NSFNET still supported network research, albeit on peering and aggregation. In addition, the NSFNET was first in pioneering research on network measurement and characterisation. As researchers became dependent on the network to support their research the ability to introduce new and risky technologies into the network became harder, as shown by the second phase router upgrade for the NSFNET when many researchers vehemently complained about any 'down time'. At this time there were still no commercial service providers from which to procure services to connect the numerous and varied sites of the NSFNET and other research networks. Hence there were still valid technical reasons for NRNs and R&E networks to exist and provide backbone services.

The policy decisions affecting the interconnection of the agency networks at the FIXes, as well as engineering international interconnectvity, were loosely co-ordinated by an ad hoc group of agency representatives called the Federal Research Internet Co-ordinating Committee (FRICC). The FRICC later morphed into the Federal Network Council in the early 1990's and finally into the large Scale Network (LSN) working group by the mid 1990s. The FNC wisely left the management of the Internet protocols to the IAB, the IETF and the IESG; however, the FNC did not completely relinquish its responsibility, as was evident by its prominent role in prodding the development of CIDR (Classless Internet Domain Routing) and originating the work that led to new network protocols (e.g. IPv6).

The Next Generation NSFNET: Early 1990s

During the early 1990s, the Internet evolved and grew larger. It could no longer remain undetected on the government policy radar screen. Many saw the NSFNET and agency networks as competing with commercial service providers (ISPs). Due to the charters of the agencies of the US based RNs (e.g. NSF, DOE, NASA), all traffic crossing their networks had to adhere to their respective Acceptable Use Policies (AUPs). These AUPs prohibited any 'commercial entity to commercial entity traffic' to use a US government supported network as transit. In addition, the demand for generic Internet support for all types of research and education communities became much stronger, and at the same time there was growing support among the US Congress and Executive Branches to end the US Federal Government support of the USA Internet backbone.

In response to these pressures and the responses to a NSF draft 'New NSFNET' proposal, the NSF elected to get out of the business of being the Internet backbone within the USA. This policy change was the nexus for the design of the vBNS, Network Access Points (NAPs),and Routing Arbiter (RA) described in the ABF paper. by early 1992. The vBNSwas meant to provide the NSF supercomputer sites a research network that was capable of providing the high end network services required by the sites for their Metacenter, as well as to provide the capability for their researchers to perform network research since the centres were still the

"Being on the net"
will change our way of doing
e-everything.

locus for network expertise. The NAPs were designed to enhance the AUP free interconnectivity of both commercial and R&E ISPs and to further evolve the interconnection of the Internet started by the Federal Internet eXchanges (FIX) and Commercial Internet eXchange (CIX). The research associated with NRNs is already evolving from dealing with mainly IP and transport protocol research to research addressing the routing and peering issues associated with a highly interconnected mesh of networks. Research was an integral part of the NAP and RA design, but it was now focused on peering of networks as opposed to the transport layer protocols themselves. Although this network was not official until 1995, commercial prototype AUP free NAPs (e.g. MAE-EAST) immediately sprang up and hastened the transition to a commercial network. The network was transformed from a hierarchical network topology to a decentralised and distributed peer to peer model. It no longer existed for the sole purpose of connecting a large aggregation of R&E users to supercomputer centres and other 'one of a kind' resources. The NAPS and the 'peering' advances associated with the NAPS were a very crucial step for the success of applications such as the world wide web (WWW) and the subsequent commercialisation of the Internet as it provided the required seamless interconnected infrastructure. Although some ISPs, e.g. UUNET and PSInet, were quickly building out their infrastructure at that time, there still existed the need for PNRNs to act as brokers for acquiring and managing end-to-end IP services for their R&E customer base; but it would not be much longer before the ISPs had the necessary infrastructure in place to do this themselves.

The Internet 2 Phase: 1996-2000

The transition to the vBNS, NAP, and RA architecture became official in the spring of 1995 and as a result the USA university community lost its government subsidised production backbone. NSF supported regionals had lost their support years earlier and many had already transitioned to become commercial service providers, an the NSF 'connections' program for tier 2 and lower schools persisted because it was felt (policy wise) that it was still valid to support such activities. The result of this set of affairs led to the creation of the Internet 2. Many of the top research universities in the USA felt that the then current set of ISPs could not affordably provide adequate end-to-end services and bandwidth for the academic community's perceived requirements. As a result, the NSF decided to again support production quality backbone network services for an elite set of research Institutions. This was clearly a policy decision by NSF that had support from the US Congress and Executive branches of government, even though in the early 1990s both Congress and the Executive Branches were fairly vocal about not supporting such a network.

The initial phase was to expand to the vBNS and connect hundreds of research universities. The vBNS again morphs from a research network, connecting a few sites and focusing on network and metacenter research, back into a production research network. The vBNS is soon eclipsed by the OC-48 Abilene network. Gigapops, which are localised evolutions of NAPS, are used to connect the top R&E institutions to the I2 backbones (i.e., vBNS and Abilene). These backbones were subject to COU as a way to restrict the traffic to that in direct support of R&E, much like the NSFNET was subject to its AUP.

The ISPs who complained so bitterly about unfair competition in the early 1990s no longer cared as they had more business than they knew how to handle in selling to corporate customers.

An ironic spin on this scenario is that the business demands placed on the commercial ISPs by the late 1990s drove them to aggressively adopt new technologies to remain competitive. Not only were they willing to act as testbeds, they paid for that privilege since it gave them a competitive edge. The result is that in a lot of cases regarding the demonstration and testing of backbone class technologies the R&E community was time wise behind the commercial sector.

This is situation is further aggravated by the fact that many, but not all, backbone network savvy R&E folks went to work in industry. Another side effect of this transition is the loss of available network monitoring data. The data used by CAIDA, NLANR, and other network monitoring researchers had been gathered at the FIXs where the majority of traffic used to pass. With the transition to a commercially dominated infrastructure, the availability of meaningful data becomes harder to obtain. In addition, as a result of the Internet 2 network's COU, and the type of applications it supports (e.g. trying to set bandwidth speed records), the traffic passing over its networks can no longer be assumed to be representative Internet data and its value in this regard is diminished.

Another milestone is reached. ISPs have grown or merged so that they are offering both wide area and local area network services and anyone can now easily acquire national and international IP and transport services. The deployment and use of virtual private networks (VPNs) allows the commercial SPs to provide and support various acceptable policy networks with differing AUP/COU on the same infrastructure. The technical need for most PNRNS or NRNS to exist to fulfill this function fades away. Researchers should now be able to specify wide area network support as a line item in their research proposal budgets, just as they do for telephony and computing support. Most governments do not support separate research POTS (plain old telephone system) networks so that researchers can talk with one another. They provide funding in the grants to allow the researchers to acquire this from the commercial sector. However, there still exist valid technical reasons for selectively supporting research networks . A prime example is the CA*Net 3 network in Canada, which has been extremely aggressive in the adoption and use of pre-production optical networking technologies and infrastructure and has been instrumental in advancing our knowledge on this area.

During this evolution of research networks capabilities, network research is also going through its own evolution. DARPA starts focusing its research on optics, wireless, mobility, and network engineering as part of its Next Generation Internet program. In addition, the research moves up the food chain of network layers. DARPA and DOE start supporting research on middleware. Globus, which along with Legion, Condor, POLDER are major middleware research efforts that become the main impetus for GRIDs; and although they are mainly focused on seeking the holy grail of distributed computing, many of the middleware services they are developing are of value in a broader research and infrastructure context. The focus of network research and research networks now starts moving away from backbone transport services to research on advanced collaboratory, ubiquitous computing, mobile, nomadic, and electronic persistent presence (EPP) environments.

The policy management of the Internet now becomes an oxymoron and reflects the completion of the transition of the Internet to a distributed commercial Internet. Many organisations are now vying for a say in how the Internet evolves. Even the IETF is suffering from its own success. It now faces many of the same political challenges the ITU faced, i.e., some commercial companies now try to affect the standards process for their own benefit by introducing standards contributions and only later disclosing the fact that they have filed patents on the technology in question. It is now much more difficult to make policy decisions regarding the future of Internet protocols, technologies and architectures.

Future Frontiers

Ubiquitous computing (UC) and electronic persistent presence (EPP) are the paradigm shift at the user level that are already drastically altering our concept and understanding of networks. The scale, number, and complexity of networks supporting these new applications will far exceed anything we have experienced or managed in the past. Users will 'be on the net' all the time, either as themselves or indirectly through agents and bots. They will be mobile and nomadic. There will be 'n' multiple instances of a user active on a network at the same time, and not necessarily from the same logical or geographical location. The frontiers associated with this new focus are many times more complex from a systems integration level than any work we have done in the past with backbone networks. This new frontier will provide new technical challenges at the periphery of the network, i.e., the intelligent access and campus networks necessary to support these new environments. EPP and UC will drastically affect out research networks and application environments, much as the WEB and its protocols drastically changed Internet and traffic patters in the 1990s.

The frontiers faced by research networks of the future will depend upon a number of technical and socio-political factors on a variety of levels. The socio-political frontiers can be broken into two different classes, one for e-developed nations who have already gone through the learning process of building an Internet based infrastructure, and another for the e-challenged nations who still face the challenges of building a viable network transport infrastructure. The developed nations need to now grapple with how they can encourage the next evolutionary phase of their Internet based economies. Due to the fast evolution of technology, the technical need for subsidising transport based network infrastructure is no longer the pressing need it was in the 1990s. The future research network will most likely be nothing more than a VPN based on a commercial ISP 'cloud' service that interconnects researchers. The High Energy Physicists (HEP) have already proven that life as a VPN based affinity group overlaid on production network services is a viable solution to providing for their network requirements. HEPNET is a virtual set of users and network experts using ESnet and other ISP VPN based network services to support the HEP scientists. Although we still have some technical challenges associated with backbone network technology (e.g., optics), there are now only a very small number of institutions and organisations capable of working with industry and making substantial contributions in this area.

The new technical challenges that need to be addressed now include how to build and deploy intelligent edge and campus networks, content delivery and routing, mobile/nomadic/wireless access to the Internet, and the support for both ubiquitous computing and electronic persistent presence. The latter two require major advancements and will require a whole bevy of middleware that is both network aware and an integral component of an intelligent network infrastructure. This includes ,but is not limited to directories, locators, presence servers, call admission control services, self configuring services, mobility, media servers, policy servers, bandwidth brokers, intrusion detection servers, accounting, authentication, and access control. IRNs and RNS can contribute to our knowledge and growth of these new areas by acting as leaders in areas that tend to more difficult for the commercial sector to address, for instance the development and deployment of advanced end-to-end services that operate over one or more ISP provided clouds. Examples include inter domain bandwidth broker services, multi PKI (public key infrastructure) trust models, defining multi site policies and schemas for directory based policy services, and developing scalable naming conventions.

In order for policy makers to make informed decisions on the evolution and support of Internet technologies and architectures, they will need to access to a generic mix of real backbone network data. There still exists a dire need at this point for such data. Innovative solutions that respect the privacy and business concerns of all types of ISPs and RNs, while at the same time making available ‘scrubbed’ data, need to be developed. In addition, with the new focus on edge and metro networks we might be able to shift our monitoring attentions to this area as well in order to better understand traffic demands and patterns on these scales of networks. Network monitoring is only one of the challenges facing us. As the scale and complexity of networks grows, even at the pico and body area network level, we will need to develop new techniques to support network modelling, simulation, and experimentation. The University of UTAH is developing a test facility comprised of a large number of networked processors, the network equivalent of a supercomputer centre, to be used experimentally in the design and development of new transport layer protocols.

Summary

‘Being on the net’ will change our way of doing e-everything and the evolution of the underlying infrastructure will need to change in order to support this paradigm shift. The intelligence of the network will not only move to the periphery, but even beyond, to the personal digital assistant and body area network. Therefore, it is important that the goals and focus of the research networks also evolve. Leave the R&D associated with backbone networks mainly with the commercial sector as this is their raison d'etre. The research networks of the future will by and large be VPNS, with a few exceptions as noted earlier in this paper. Research networks need ot focus on the new technologies at the periphery as well as the middleware necessary to support the advanced environments that will soon be commonplace. Many research networks will themselves become virtual, e.g., HEPNET, providing expertise but not necessarily a network service. Policy makers must adapt to address not only these substantial technical and architectural changes but they will also need to deal with second order policy issues such as security and privacy and how to ensure that we don't end up with a bifurcated digital economy of e-savvy and e-challenged communities.

E-Developed nations have already been through the technology learning curve of implementing and deploying a transport infrastructure. The e-challenged nations, with respect to network infrastructure, still face these same challenges and they have the benefit of leveraging the knowledge of the nations who have successfully made the transition. In order to speed up the deployment of Internet technologies and infrastructure in the e-challenged nations, it may be best to first create technologically educated people and then to provide them an economic and social environment where they can apply their knowledge and build the infrastructure. E-savvy nations should help by providing the ‘know how’. NATO has a joint program with TERENA to provide for the instruction of Eastern European nations on the use and deployment of Internet technology (i.e., how to configure and manage routers). In lieu of subsidising networks in these nations, NATO and TERENA are providing the basic knowledge that these people need to build, manage, and evolve their own networks and infrastructure. This should be the model to consider for e-developing nations. This is not to diminish the challenges of building network infrastructure in some areas where there is no such infrastructure, and perhaps in some of these areas working with other utility infrastructure providers might advance this cause.

Notes

1 This is also attributed to the famous Physicist Niels Bohr.

2 Wulf, William A. 1988. The national collaboratory – A white paper, Appendix A. In Towards a national collaboratory. Unpublished report of a National Science Foundation invitational workshop. Rockefeller University, New York. March 17-18, 1989

3 http://www.nsf.gov/

4 http://www.darpa.mil/

5 http://www.gigaport.nl/

6 Draft-aiken-middleware-reqndef-01.txt, IETF RFC, May 1999, http://www.anl.gov/ECT/Public/research/morphnet.html

7 See http://www.dante.org/ and http://www.terena.nl/ for full lists of European research networks.

8 http://www.nordu.net/

9 http://www.canarie.ca/

10 http://www.internet2.org/

11 ‘Architecture of the Multi-Modal Organizational research and Production Heterogeneous Network (MORPHnet)’, Aiken, et al, ANL-97/1 technical report, and 1997 Intelligent Network and Intellgence in Networks Conference. http://moat.nlanr.net/Papers/iinren.ps

High Energy Physics Data Grid Initiative

This is a summary of part of a presentation made by Federico Ruggieri, chairman of the HEPCCC, at the initial HEP Grid meeting at CERN 11 January 2000

The conclusion of a discussion at the HEPCCC meeting on 12 November with senior representatives of the European Commission's IT programme (George Metakides and Thierry van der Pyl) was that a proposal from HEP for a project concerning the development of a Computational and Data Grid would be considered seriously. The Grid concept, introduced in the recent book by Foster & Kesselman, provides a metaphor for a coherent set of computing resources physically distributed across a number of geographically separate sites. A Computing Grid exhibits a uniform interface to its resources, providing a dependable service which can be accessed from anywhere. This can in some respects be compared to the electric power grid. A number of Grid projects have been initiated or proposed in North America, in particular the Globus project, which has developed a toolkit implementing a set of Grid services within a layered architecture. A number of European companies and organisations has started the E-Grid Forum http://www.egrid.org, paralleling the American organisation http://www.gridforum.org.

An initial proposal for a HEP EU-Grid Project has been made Federico Ruggieri, which would focus on:

Exploiting the Grid concept and technology in the field of HEP; Deploying a large scale implementation of a computational and data Grid using technology developed by existing projects, complemented by the middleware and tools necessary for the data-intensive applications of HEP.

The overall topology of the Grid would follow that of the MONARC Regional Centre model, with a number of national grids (equivalent to the Tier 1 and Tier 2 Regional Centres) interconnected by a central node at CERN.

The main target of the project would be the HEP community, but it would be open also to other scientific communities. The necessary high bandwidth networking infrastructure would be funded by an independent activity, such as the EU supported GEANT (networking) project. The project would collaborate closely with similar North American initiatives such as the Particle Physics Data Grid. The partners should be institutes, funding agencies and industrial companies, with CERN as the lead partner and coordinator. A steering committee would drive and guide the project, including representatives of the LHC experiments in addition to the operating partners.

The DataGrid Project

The DataGrid Project is a proposal made to the European Commission for shared cost research and technological development funding.

The project has six main partners:

CERN The European Organization for Nuclear Research near Geneva on the French/Swiss border
CNRS France – Le Comité National de la Recherche Scientifique
ESRIN The European Space Agency's Centre in Frascati (near Rome), Italy
INFN Italy – Istituto Nazionale di Fisica Nucleare
NIKHEF The Dutch National Institute for Nuclear Physics and High Energy Physics, in Amsterdam
PPARC United Kingdom – Particle Physics and Astronomy Research Council

and fifteen associated partners:

CESNET Czech Republic
COMMISSARIAT À L'ENERGIE ATOMIQUE (CEA) France
COMPAGNIE DES SIGNAUX Systémes d'information – France
COMPUTER AND AUTOMATION RESEARCH INSTITUTE Hungarian Academy of Sciences (MTA SZTAKI)
CONSIGLIO NAZIONALE DELLE RICERCHE (CNR) Italy
DATAMAT Ingegneria dei Sistemi S.p.A. – Italy
HELSINKI INSTITUTE OF PHYSICS Finland
IBM UNITED KINGDOM LIMITED
INSTITUT DE FISICA D'ALTES ENERGIES (IFAE) Barcelona – Spain
ISTITUTO TRENTINO DI CULTURA (IRST) Italy
KONRAD-ZUSE-ZENTRUM FÜR INFORMATIONSTECHNIK Berlin – Germany
ROYAL NETHERLANDS METEOROLOGICAL INSTITUTE (KNMI)
RUPRECHT-KARLS-UNIVERSITÄT HEIDELBERG Germany
STICHTING ACADEMISCH REKENCENTRUM AMSTERDAM (SARA) Netherlands
SWEDISH NATURAL SCIENCE RESEARCH COUNCIL (NFR) Sweden

The objective of the project is to enable next generation scientific exploration which requires intensive computation and analysis of shared large-scale databases, from hundreds of TeraBytes to PetaBytes, across widely distributed scientific communities. We see these requirements emerging in many scientific disciplines, including physics, biology, and earth sciences. Such sharing is made complicated by the distributed nature of the resources to be used, the distributed nature of the communities, the size of the databases and the limited network bandwidth available. To address these problems we propose to build on emerging computational Grid technologies, such as that developed by the Globus Project to:

- establish a Research Network that will enable the development of the technology components essential for the implementation of a new worldwide Data Grid on a scale not previously attempted;
- demonstrate the effectiveness of this new technology through the large-scale deployment of end-to-end application experiments involving real users;
- demonstrate the ability to build, connect and effectively manage large general-purpose, data intensive computer clusters constructed from low-cost commodity components.

These goals are ambitious. However, by leveraging recent research results from and collaborating with other related Grid activities throughout the world, this project can focus on developments in the areas most affected by data organisation and management.

The three major thrusts of this activity are:

- computing fabric management, including network infrastructure, local computing fabric (cluster) management, and mass storage management;
- data grid services to provide workload scheduling, data movement, and Grid-level monitoring services;
- technology demonstration and evaluation using scientific applications in three major disciplines – high energy physics, earth observation, biology.

The structure of the programme of work is as follows:

Five sub-projects (work packages): Grid Workload Management, Grid Data Management, Grid Monitoring Services, Fabric Management, and Mass Storage Management, will each develop specific well defined parts of the Grid middleware.

The Testbed & Network activities will integrate the middleware into a production quality infrastructure linking several major laboratories spread across Europe, providing a large scale testbed for scientific applications.

Three scientific work packages – High Energy Physics, Earth Observation Sciences, and Biology – will adapt existing applications to use the Data Grid software, and demonstrate them operating on the successive releases of the testbed – with a focus on testing the quality as well as the functionality and performance of the system.

The IBM Blue Gene Project

Marc Snir
IBM T.J. Watson Research Centre
Oct 13, 2000

The Blue Gene Project Mission

On December 6, 1999, IBM announced a $100 million research initiative to build the world's fastest supercomputer, 'Blue Gene', to tackle fundamental problems in computational biology. The Blue Gene system will be capable of performing more than one petaop/s (1,000,000,000,000,000 operations per second). The Blue Gene project will use this computer for large scale biomolecular simulation to advance our understanding of biologically important processes, in particular our understanding of the mechanisms behind protein folding. The project will also advance our knowledge of cellular architectures (massively parallel systems built of single chip cells that integrate processors, memory and communication), and of the software needed to exploit those effectively, for a variety of applications.

Blue Gene is Long term Research

IBM Research always has a mix of short term and long term activities. Indeed the same research project will often start as a long term activity, unrelated to any existing IBM product and mature, over time, into technology that is relevant for an existing or planned IBM product. Both types of activity are essential to IBM Research. Long term research is necessary to fill the technology pipeline of IBM; to maintain the vitality of the IBM research community; and to attract top researchers into IBM. Short term research is necessary for the IBM Research Division to be relevant to IBM; and for the IBM research community to be aware of the problems and needs of the Computer Industry.

Like any ambitious long term research project Blue Gene is a complex, high-risk project that addresses fundamental, challenging research questions. In computational biology, we need to understand to what extent numerical simulation can shed light on the real-life protein folding mechanisms. Since simulations on this scale are unprecedented, we may discover that better numerical models are needed. In computer science, we need to understand how one can programme a machine of this size. Since no-one has attempted parallelism at this scale, we may discover that new programming models are needed.

The mode of operation on long term projects is necessarily different than on short term projects. Long term projects are more science oriented, and have more of an academic flavour. They aim more at producing basic knowledge, less at producing reusable technology. They are higher risk than short term projects. They are typically more open, with a freer flow of information between IBM and the academic world, and with more significant collaborations with the academic world. Collaborations with the external research community will be key to addressing the research challenges of the Blue Gene project and in deciding how to best utilise this powerful computational resource in order to advance science. We expect that information about the Blue Gene research will be widely available and freely published as would be information about a similar academic research project. This, because the advantage of better interactions with academia more than compensate for the potential disadvantage of loosing control over proprietary information.

The Blue Gene project is a 'big science' project: it involves a large team of people with very diversified skills; logic designers, packaging and cooling specialists, microprocessor designers, firmware and system software specialists, compiler, libraries, and programming tools specialists, algorithm and numerical methods specialists, computational chemists and biologists. This team will work in close collaboration over multiple years on the Blue Gene project. IBM Research is uniquely positioned to undertake such big science projects. It is perhaps the only research lab in the world where skills in all relevant areas can be found. And it is an organisation that can afford to take the long view and provide a stable environment for big, ambitious, multi-year projects. On the other hand, IBM Research has much less of an advantage over academia for 'small science' projects that can be completed in one or two years by a small team of few people.

Expected Blue Gene Project Outcomes

Cellular Computer Architectures Long-term projects at IBM Research are expected to attack problems that are fundamental to the long term evolution of the computer industry. Blue Gene is no exception. The continued evolution of the silicon technology has brought us to a point where one can integrate a full system – processor, memory and I/O on a chip. This is due both to the continued growth in the number of transistors per chip, and to the emergence of CMOS technology that supports well both embedded DRAM memory and logic on the same chip. This capability can be used to solve fundamental problems in the design of current microprocessors and large servers.

VON NEUMANN BOTTLENECK While microprocessor logic speed has increased rapidly in the last decades, memory access time has barely decreased. As a result, the relative memory access time has worsened. A modern microprocessor might execute many hundreds of operations in the time it takes to load data from memory. If the memory latency problem was not addressed then a microprocessor would idle most of the time. Indeed, most of the logic in a conventional microprocessor is dedicated to the hiding of memory latency. The problem almost disappears if the memory is on the same chip as the microprocessor. This is the approach followed by Blue Gene.

MICROPROCESSOR COMPLEXITY The main impetus in microprocessor design is that of achieving best possible performance on one sequential thread of execution. To that purpose, microprocessors have become increasingly complex, with greatly diminishing returns on added logic complexity. Since the advent of single chip microprocessors, the number of gates per microprocessor has increased by three order of magnitudes, and the total design time has increased by two orders of magnitude. Yet, the number of instructions executed per cycle has increased by less than an order of magnitude.

This choice is rational for most microprocessor uses However, it is highly irrational for applications that are highly parallel: one can get a much better throughput for such applications by using a larger number of simpler execution units, thus avoiding the diminishing returns of modern micro architectures. Such a move has many additional advantages: The design is simplified to the extent that a small design team can realise the system. The power consumption for a given level of performance is reduced. This is very important, since the top performance of large supercomputers is limited today by their power requirements, which is measured in Megawatts. The physical size of a microprocessor is reduced. One can cover a chip with an array of simple, slower microprocessors, rather than a more complex, but faster microprocessor. This is important as it is hard to communicate to the entire width of a chip at current logic speeds. Furthermore, it becomes possible to use chips even if parts of the chip are faulty, using the working microprocessors. This can significantly improve yields and thus reduce costs of microprocessor chips.

PROCESSOR-MEMORY BALANCE The Blue Gene system will achieve x100 more performance than a conventional supercomputer built from a similar number of chips. This is due to the much higher level of integration in Blue Gene, but also to the fact that Blue Gene will have less memory than such a supercomputer, notwithstanding its much higher performance. Blue Gene will be capable of execution more than a petaop/s, but will have only a fraction of terabyte (1,000,000,000,000 bytes) of memory. This would seem to indicate that Blue Gene is an ‘unbalanced’ system, where it will not be always possible to utilise the execution units efficiently. But is this bad? Only a very small fraction (few percent) of the circuits in a conventional system are used to actually perform arithmetic or logic operations. Most circuitry in such a system is dedicated to communication, i.e., moving data from one part of the system to another, and to storage, i.e., moving data in time. Given this state of affairs, it seems logical to increase the number of execution units in the system, We need to understand to what numerical simulation can shed light on the real-life protein

even this decreases their utilisation, provided that storage or communication logic is better utilised. Blue Gene can be thought of as an 'active memory' system, a large array of memory chips with processors embedded in the memory chips. Algorithm design can focus on reducing communication and storage use, since operands can be processed wherever available in the system, as soon as available. We expect that a better understanding of the real trade-offs between resources in current silicon technology will lead to new algorithms and new computation methods that can achieve better overall performance from a given silicon budget, even if they do so at the expense of a much higher operation count. Our initial study of the molecular dynamic algorithms used for protein folding indicate a much better overall performance for algorithms that have a higher operation count, but have a simpler communication structure. Furthermore, the memory requirements for these algorithms is a small fraction of the total (small) amount of memory available in Blue Gene. Similarly, this huge machine will generate, for molecular dynamic simulations, an I/O stream that will be easily satisfied by a few hundreds of disks.

ERROR HANDLING Some computations on Blue Gene will have to go on for weeks, if not months. On the other hand, on a machine this size, hardware failures will be frequent enough to bring down parts of the system every few days. Furthermore, there will be a non negligible chance that a long running computation will have been corrupted by an undetected hardware error. This forces us to address very seriously the problem of error handling in Blue Gene.

The Blue Gene system will be built so that any component can be isolated while the entire machine continues to run. Furthermore, the Blue Gene software will be designed so that applications can be restarted from a checkpoint and remapped automatically onto a partially available Blue Gene. This will take care of detected errors. Errors that are not detected by hardware can be handled simply by mirroring – computing everything twice and comparing. But this effectively halves the machine performance. Less onerous software checks can be used in many cases to reduce the likelihood of undetected errors to an acceptable threshold. The issue of error handing is essential for future computer system, as the decreasing size of transistors increases the likelihood that a transistor 'misfires', while the increasing number of transistors in large systems magnifies the effect of such errors.

SOFTWARE FOR MASSIVELY PARALLEL SYSTEMS While theoreticians have thought of computers with millions of processors for many decades, it is only now that it becomes feasible to build such systems. Furthermore, if, as many predict, the rate at which silicon technology improves slows down in the coming years, then massive parallelism may be the only way to reach certain levels of performance in the future. But the design of massively parallel systems, even if feasible from the hardware side, raises very hard questions on the software side: which problems can be solved efficiently on a massively parallel machine? How should such machines coded? How can one control such machines? How can one isolate problems when they occur, looking for the proverbial needle in the haystack? All these issues, and many more, require a vigorous research program on software for massively parallel systems. Blue Gene is the first system where such research can be applied.

Research in Computational Biology The Blue Gene machine will provide an opportunity to do top quality research in computational biology. A key research area where Blue Gene will be applied is protein folding. Proteins are simply chains of amino acids of which there are 20. The sequence of amino acids in each protein in our body is encoded in the DNA genetic code. Proteins in a water solution fold into a unique intricate three dimensional pattern. The function of a protein strongly depend on this three dimensional structure. Hence, to understand what function is coded by a DNA sequence that defines a protein, one needs to determine how is the three dimensional structure of a protein determined by its sequence of amino acids. A simple, if extremely expensive method of doing this is to simulate the process of protein folding. This is done by developing a classical, atomistic model of protein water interaction, where atoms are assumed to move as a result of interatomic forces acting upon them: e.g., long range electrostatic forces, or forces that approximate the interaction between bonded atoms. A computer can now be used to simulate the evolution of this mechanical system, using discrete time steps. Unfortunately, accurate simulations require time steps of few femtoseconds (10^{-15} second), while a molecule may require milliseconds, or even seconds to fold. Thus, such a simulation requires many time steps, hence an extremely powerful machine.

It is not obvious that such a 'brute force' simulation will work. The classical models used are, after all, approximate, and may not provide enough accuracy for long term simulations. Also, other methods, that are less compute intensive, are likely to be more effective in unravelling parts of the 'protein folding problem"; one can often 'guess' the final folded structure of a protein without simulating in detail the evolution of the folding process. However, it is very likely that the ability to perform simulations at a scale hitherto unachievable will significantly advance our understanding of the folding process and our ability to simulate biomolecular systems. Advances in our understanding of biomolecular simulation are relevant to a variety of related problems such as drug protein interactions or enzyme catalysis.

Computational Biology research is in an area where future technology is important to IBM. Computers are increasingly used in biological research. The hope is that skilled use of the mass of genomic information that is becoming available and skilled use of computer simulation will significantly accelerate the development of new drugs. This quickly growing, multibillion market is one that interests IBM very much. Basic research at one end of the pipeline is needed in order to have at the other end of the pipeline the know-how, the insight and the technology required for products in this market.

Open-Source Software for Research

Philippe Aigrain
Head of Sector 'Software Technologies'
Information Society Technologies programme, European Commission, DG INFSO/E2

DISCLAIMER This communication represents only the personal opinion of its author, and does not necessarily reflect the view of the European Commission

Software as embodiment and enabler of research

The importance of open-source software for global research cannot be presented without first mentioning the essential role that software in general plays in today's scientific research. The research infrastructure from equipment to documentation, from networks to journals is powered by software. Whether in astronomy, chemistry or biology, in material sciences or in environment and climate science, the scientific results themselves are more and more often embodied in software, or data and models closely integrated with database and simulation software.

Open-source software: definition

In this context, the intellectual property and licensing schemes that apply to software have an essential impact on science production and dissemination. Open-source software (originally termed free software[1]) is software released under a licensing scheme[2] authorising users to freely access the source code, modify it, compile it, use the resulting executable and redistribute the possibly modified code. On the basis of such licensing schemes, a full universe of software (from editors to compilers, from operating systems to desktops, from databases to application software) has been built by more than a hundred thousand contributors[3]. A development model has also emerged, based on distribution, co-operation, code review, criticism and competition within co-operation.

Open-source software: the scientific method at work

In addition to being a major opportunity for global science, open-source software is closely related to the scientific method itself, with which it shares important features: visibility of internal principles, open exchanges, criticism, co-operation, cumulative build-up, importance of esteem of colleagues in rewards to contributors.

Co-operation and cumulative knowledge: co-evolution with the Internet and the Web

The possibility and the success of world-wide distributed co-operation in software and science would not have materialised without the technical infrastructure of the Internet and the World Wide Web, of which many critical components are themselves based on open-source software. Easy inter-personal communication and global information publishing at low cost are the key enablers of co-operative development. The future development of open-source software calls for maintaining the integrity and the openness of these infrastructures.

Open-source software as a scheme for publicly funded research results

Open-source software first bears the promise of delivering public results for publicly funded research. Many think that the dissemination and re-use of research results embodied in software, in particular those obtained in co-operative research with strong industry participation, have greatly suffered from restrictive intellectual property licensing schemes. The transition to a more open framework of licensing will not be a smooth one, as it calls for a mental revolution in selection criteria and contracting frameworks, technology transfer mechanisms, links between research and business models. The transition is of particular importance in fields like biotechnology and information and communication technology itself.

ACCESS TO ADVANCED ICT INFRASTRUCTURES

SHARING OF DATA, INFORMATION, RESOURCES AND FACILITIES FOR RESEARCH WITH ICT

KNOWLEDGE DISSEMINATION AND ACCESS TO KNOWLEDGE

software lowers entry barriers for science innovators.

Lowering entry barriers for science in developing countries

One of the most appealing benefits of open-source software (particularly if it is combined with open information) is the lowering of entry barriers for science innovators in countries with limited resources. This also applies in regions of developed countries and to smaller entities in innovation. Of course, this lowering does not magically cancel all obstacles to taking part in the global research village: insufficient basic education, the required physical equipment, the cost of travel, the lack of everyday face-to-face exchanges with other creative minds in the same field, or simply the living conditions are still important obstacles.

Challenges and possible limits of the scheme

Open-source software still has a way to go in terms of developing usability, and creating confidence on the availability of support services by companies. In addition, its success itself exposes it to new challenges. The ever-increasing number of applications makes it difficult to locate the right one. There is some inflation of functionality in each application, though to a lower extent than in proprietary software. The complexity of supporting an ever increasing variety of devices and communication channels is a challenge. Though the open-source development model offers powerful mechanisms to deal with these problems, they will represent difficult challenges in particular for configuration and software architecture management.

The business models associated with open-source software are service-based. This can take many forms. In one scenario, one would witness a rich variety of local support, training, additional development, project hosting or packaging and distribution services. In other scenarios, the same concentration trends that are powerful in packaged software would also develop, and result in new types of imbalance between players and between countries.

Practical implementation

In addition to the promotion of open-source software as a scheme for dissemination of publicly funded research results, the support to open-source software as a basis of global science is emerging under the following axis:

- Use in scientific education
- Commitment to open data standards and formats, with open-source software reference implementations
- Creation of repositories of software, in particular related to one scientific field, and often with associated free information resources
- Support to networking infrastructure, software engineering tools and support services for distributed development[4] and virtual laboratories[5]
- Development of future open-source based information infrastructure, for instance in grid computing
- Support to the development of open-source software in scientific or technical 'niches' (space, high-performance computing), if one is able to create a sufficient base of users/developers

Notes

1 The Latin word *libre* avoids the ambiguity between freedom and free-of-charge. Thus, in Latin countries, the original expression 'logiciels libres' or 'software libre' is still used preferentially.

2 See for instance http://www.fsf.org/philosophy/license-list.html

3 This figure is based on a number of independent estimates using databases of free software contribution and registered users of software development hosting services.

4 See actions developed in the IST Work Programme 2000 and 200, http://www.cordis.lu/ist/ka4/tesss/impl_free.htm

5 See for instance http://www.unesco.org/webworld/build_info/networks.html

The Power of Distributed Computing

Dr. Larry Smarr (Interview)
Professor of Computer Science and Engineering, University of California at San Diego
Director Designate, California Institute for Telecommunications and Information Technology
Strategic Advisor, National Computational Science Alliance

The very nature of technological prevents accurate predictions for periods longer than about five years. However, there are some trends that I think will have a major impact on scientific computing, communication and access to knowledge. A crucial element is the continuing trend towards faster and cheaper computers. Today's Pentium III desktop computers are as powerful as supercomputers, for example a Cray Y-MP of ten years ago. The speed creates a competitive advantage. By being able to handle a larger number of variables, much more complex problems can be tackled. Because of the trend towards commodity components in high-end computing, undreamed of computational capability will be available at very low cost for any scientist. The computer is becoming THE most important scientific tool in a growing number of scientific fields. In biology, for example, Celera was able to sequence some parts of the human genome earlier than the NIH efforts because they used a far more computer intensive approach. It is also true for the social sciences, because the greater speed and memory of computers enables one to tackle large dataset problems enabling a more realistic model of the true complexity of a social environment.. Of course, a crucial problem for the social scientist is the availability and quality of data, an important but often neglected point. The collecting of data is often not funded by governments, the US government is especially weak in this respect due to the political sensitivity of much social science research.

Scientists will increasingly use the Web as a general interface to a fully digital working environment. We are witnessing the emergence of a digital grid in which all components of our research and education workplaces are merged seamlessly. Logically, we can distinguish two different type of grids: the *computational* grid and the *access* grid. Through the computational grid scientists will have access to virtually unlimited computing and distributed data resources. The access grid will create a group collaboration environment. In NSF's Partnerships for Advanced Computational Infrastructure program, researchers are developing middleware designated to make complex grids manageable. Such software should make it possible for Grid-enabled applications to independently discover available network, computing and storage resources and to obtain advance reservations for such resources. Security and encryption issues, crucially important in an era of distributed computing, will also be handled by such middleware tools.

This new development to Web based distributed computing is an extension of the development of computer architecture during the last couple of decades. Every five years, we have seen a new machine develop and every ten to fifteen years a completely new way of thinking in high-end computing emerged. The vector machine was invented 25 years ago and it was dominant for fifteen years. Then, the expensive custom-made vector processors were displaced by RISC workstations and later by Intel PC processors. This development has created a completely new economic rationale in high-end and supercomputing. The greater improvement in the speed of PC processors accelerated the new economy in computing. By using clusters of PCs, scientists no longer needed to to purchase expensive commercial supercomputers for much of their work. For most complex scientific problems, networks of ordinary personal computers turned out to be good enough.

Now, a new change is coming up: internet computing. Why stop at clustering a few hundred PC's if you can combine millions of PCs through the Web, creating megacomputers? Computing will become a distributed scalable function from the internet megacomputers, to thousands of processors in national centres, hundreds at universities, dozens in laboratories and one of each four desktops working part-time on computational problems posted on the Web. A very nice example of this trend is the SETI@home project which carries out signal processing looking for alien life forms broadcasts. To process all these signals from outer space and try to discover meaningful patterns in them, they have spread out the computations over more than 2.5 million computers via the Web. A number of start-up companies, such as Entropia in San Diego, are commercializing this approach to create an internet based computer fabric.

This will allow the Web to become self-powered with intelligent agents drawing computing power from PC's all over the Web. Computing power will become organised like the electricity grid. This new megacomputer is unlike any existing in one university or country. We will get a global supply of computing power. Private companies will provide access, like oil is being supplied now, but it will be globally accessible. Companies will specialize in optimizing this for specific fields or problem areas. I think this is an inevitable development. Distributed computing becomes the normal way of working. It will be applicable to a wide range of scientific problems: bio-informatics, cancer and AIDS research, climate research, fluid dynamics and so on. This is THE groundbreaking development for the next five years in high-end computing as well as a huge challenge. It will be necessary to in some cases to rewrite the code of applications to enable distributed computing, in others we will be able to use software wrappers to hide the legacy software. Rethinking algorithms will allow innovators to recast the problems of science in these distributed terms. Not a small endeavour.

I do not foresee a large regulatory role for national governments with respect to distributed computing. So far, the US government has stayed out of the development of the internet and this has proven to be wise. Governments which try to regulate the internet and access to the internet will harm themselves economically.

This raises the problem of skills, of course, but in this respect the Web itself may provide a solution. It enables lifetime learning in ways not possible before in the form of all sorts of online training programmes. The private sector is quickly taking up these issues, much faster than the universities. I think the public sector is simply not fast enough to be able to compete in creating general frameworks for distribution, but the universities could be very important partners in providing content. For this sector to take off, it is critically important that universities and national governments play the role of early adopter and customer for such online education and training.

A serious problem in all this is the potential disjunction between the evolution of hardware and of software. High-end and supercomputing hardware is developing very fast, perhaps even faster than Moore's Law. The fastest machines currently under construction are well into the terascale range, at about ten to thirty teraflops, and petaflop machines are under consideration. The industry manufacturing these highly specialised supercomputers has been going through a restructuring process, related to the developments mentioned above, and it is not clear that there will be sufficient resources to develop efficient software for the new machines. This would seriously hamper effective use of these machines for scientific research.

Therefore, I am quite happy with the new report on open source software prepared by the Libre Software Working Group created at the initiative of the Information Society Directorate of the European Commission (http://eu.conecta.it/)[1]. Although the report covers software in general, and does not address specific problems related to scientific software, it is very valuable. It clearly spells out the nature of open source software, its contractual variations and possible disadvantages of using it. Open Source software is by the way something different from software for which one does not have to pay. Its basic feature is not that it is free in that sense, but that the source code, written in some high-level programming language, is made publicly available under such conditions that people are encouraged to test the software, develop it further and modify it. This often results in a good balance between maximum reliability and maximum creativity.

I found this report to be the best single source on the planet on this issue as I was co-authoring the Recommendations of the Panel on Open Source Software For High End Computing. This Panel was under the auspices of the United States' President's Information Technology Advisory. The European Union has been

Journals are not
where the interesting action is.

more aggressive in open source approaches than the US government. The US is still formulating their policy. Not many EU reports have influence here in the States, but this one surely will have impact. The whole issue of open source software strikes at the heart of intellectual property. The key issue is licensing, about which different views are being held by different parties. It is a very interesting unfolding experiment and I think we should let the marketplace decide which option is the best for the future.

In scientific research, one can distinguish three major methodologies: theoretical (formulating equations and analyzing), experimental/observational, and since the late 40s computational science/simulating. Von Neumann formally defined the latter approach. His major point was that the correct mathematical approach is the transformation of partial differential equations into discrete equations in order to be able to solve them. These discrete equations can then be mapped to the memory of a digital computer. This is the basic epistemology of computational science. It is widely applicable.

A good illustration of the power of the discretized approach, is the work of Kelvin Droegeimeier at the University of Oklahoma. He was able to use a large memory Silicon Graphics supercomputer at the National Center for Supercomputing Applications to predict a tornado's development faster than in real time, which is quite a feat. The main point is that to predict a small feature like a tornado one needs a grid that is more fine grained than the 30 or so kilometer meteorologists use for production national weather forecasts. Droegeimeier has developed a new technique, adaptive grids, which enable the computer to zoom in on an area where smaller scales are important and couple the finer grid seamlessly into the coarser grid. This is very useful because the weather system develops at different scales.

Weather is unpredictable for long periods of time because a very small perturbation can cause large effects. Therefore it is not enough to know the initial data, which observationally is sampled only coarsely, but one must consider an ensemble of different initial data to get 'error bars' on the final forecast. Droegeimeier takes the initial data gathered by balloons, satellites, surface measurements, etc. and then perturbs them to get not one, but a whole set of initial conditions that are all compatible with the observational data. Then he takes 20 sets of them and runs them all in real time. On top of that, he has developed a smart algorithm to add the outcomes together, which is called 'ensemble supercomputing'. The result is better than today's weatherman can produce.

This is also applicable in solving other types of complex problems. Ensemble supercomputing is very useful for any kind of environmental simulations. It takes into account the uncertainties in the system to produce results with the uncertainty quantified locally. I think these adaptive grids are the next big phase in supercomputing software algorithms. One can also look at it from the number of dimensions computers can handle. First we had computers which could tackle one-dimensional problems. Then we got to two-dimensional flow problems and in the 1990s we have been able to tackle three-dimensional simulations. However, this is still restricted to one space-scale at a time. Instead, we would like to be able to resolve all three dimensions at different spatial scales simultaneously. Adaptive grids allow the computer to do exactly this, by solving the problems automatically, speedier than a human intervention to regrid by hand can accomplish. Although in computer science this important development is well understood, most computational scientists are not yet aware of these possibilities.

An important correlated emerging area is scientific visualization. The idea behind this is that the computers nowadays generate vastly too many numbers to comprehend and interpret numerically. Therefore, the natural way of human beings, looking at phenomena, should be mobilised.

By now one can see that there are so many developments that teaming up between computational scientists and computer scientists is becoming crucial. Computers have become too complex for a scientist to really take advantage of their full potential.. So you need, in physics for example, a variety of physicists of different backgrounds, plus computer scientists, plus software engineers. This creating of new scientific teams is essential. I expect that it will make more disciplines comparable to the style of research in high energy physics. Information and communication technologies can play a large role in promoting this formation of new types of scientific teams. In this, the access grid mentioned above is crucial.

Scientific instruments, visualizations, simulations and virtual realities, scientific publications, pre-prints, colleagues elsewhere in the world: they will become accessible through commodity Web browsers. Collaboration may be speeded up because electronic, multi-media notebooks can record all these transactions. Because broadband wireless connections will become more available to scientists, one can have one's workspace whereever you happen to be.

NSF's Partnerships for Advanced Computational Infrastructure Programme is developing the idea of the Web based technical software workbench on the basis of the 'portal' concept developed in the commercial internet world. A key feature of portals is that users can customize both the content of information and the way it is being presented. They moreover have tools to create new information. The Science Portals will be based on the emerging standards for distributed computing component systems, such as Corba and Java. They will also exploit the XML metadata standards. The major difference between science portals and commercial ones will be the integration of advanced GRID services, very large scale data analysis, advanced visualization and collaboration services, and distributed components that run on high-end parallel systems. Electronic communities sustained by these technologies will become the norm in science and scholarship because at a certain point in time everybody will see the advantages of sharing in a networked world.

Notes

1 Free Software / Open Source: Information Society Opportunities for Europe? http://eu.conecta.it/).

Modelling for Knowledge Exchange: Global Aspects of Software for Science and Mathematical Modelling

Andrzej P. Wierzbicki
National Institute of Telecommunications

Marek Makowski
International Institute for Applied Systems Analysis

Abstract

The knowledge-based economy and information societies are defined similarly by knowledge and information becoming an essential or even dominant productive factor. In order to reflect on their impact on global science, it is essential to better understand the distinction between information and knowledge. Various types of understanding of the concept of knowledge are discussed in this paper, but one which is characteristic of hard sciences in the information age – is related to synthesising information into mathematical models that can be analysed by using computers. Data mining in very large data sets is also related to finding patterns or models that synthesise characteristics of data relevant for a given purpose. Most of the methodological conclusions related to mathematical modelling and data mining are not restricted to computer science, but have an interdisciplinary character and support interdisciplinary research.

Knowledge and information become either more commercialised in a knowledge-based economy – or, if supported by public funding, more accessible for public use. Thus, it also becomes more important to make more knowledge accessible in the form of mathematical models used by various scientific disciplines, or in the form of patterns derived from large data sets. Easy exchange of computerised mathematical models will help in better verification and validation of research results for a knowledge-based economy and information society. However, in order to increase such accessibility, better standards and software tools for the analysis of mathematical models should be developed. The development of such standards and tools should become one of the priorities of science policy.

1 Information civilisation: megatrends and challenges

The global information infrastructure, the information society and the knowledge-based economy will be considered here jointly as the *information civilisation*, since their main aspect is information and knowledge becoming a dominant aspect of economic activities. More precisely, *information civilisation* is understood here to be the era of the development of the *information society* or – practically equivalent – of *knowledge economy* (because a common definition of both these concepts stresses the dependence of this society and economy on information and knowledge as the increasingly dominant production factor). Under the term *megatrends* we understand here important tendencies in social or technical development that can be observed and predicted for a longer time horizon.

It is important to note that there are reasons to forecast the length of the period of information civilisation for many future decades, perhaps for the entire twenty-first century. The megatrends discussed below will also last for several decades. The reasons for these long-lasting megatrends are as follows. Many new ideas and developments of contemporary science and technology are not being implemented as fast as would seem to be indicated by the rapid progress of science from a purely technological point of view. Delays in their implementation are a result of diverse social and economic factors. One example of this is the development of digital television. Its theoretical foundations were laid over 40 years ago – and its wide implementation has not yet been fully achieved. If technological reasons were the only decisive factor, then the time required to implement digital television could have been shortened to 20 years. Social and economic reasons were responsible for much longer implementation time. There are many other such examples. For these reasons, today we can forecast rather precisely – in qualitative terms – which megatrends will be decisive for the future development of information civilisation; the uncertainty stems from the scale and timing of full implementation as well as future technological details.

The selection of trends that should be recognised as most important and relevant depends on the evaluation of their economic, social, technological and intellectual impacts. From this perspective, three basic megatrends can be selected (see Wierzbicki, 1999) as being decisive for the development of the information society and as generators of many resultant trends of more specific character:

- **TECHNOLOGICAL MEGATREND OF INTEGRATION OR CONVERGENCE,**
- **SOCIAL MEGATREND OF CHANGING AND CREATING NEW PROFESSIONS,**
- **INTELLECTUAL MEGATREND OF CONCEPTUAL CHALLENGES.**

The **MEGATREND OF TECHNOLOGICAL INTEGRATION OR CONVERGENCE** concerns media, methods and systems of transmitting and processing information, but also the practical use of information. It starts with a general *digitalisation* of the methods and systems transmitting and processing information, but includes also the trends of:

a *multimedia communication,*

b *mobile telecommunication,*

c *fast increase of transmission rates* in telecommunication networks,

d *integration of new telematic services* into complex service systems,

e *e-commerce, e-banking,* etc.

The last trend illustrates also *two facets* of this technological megatrend. While the megatrend is based on exploiting the technological advantages and possibilities of integration provided by the digitalisation of all information, it is also driven by market gains and by commercial applications. The same will concern the development of knowledge-based economy: while being enabled by technological integration, it will be driven by gains obtained from commercialisation of knowledge exchange.

Concerning the **MEGATREND OF CHANGING AND CREATING NEW PROFESSIONS** we observe that the development of the information civilisation relies on a substitution of old professions by their newer versions or by entirely new professions. Old professions require much physical effort and are badly equipped with information technology tools; new professions require much knowledge and information and utilise increasingly more information technology tools. This implies an increasing *dematerialisation of work.* There are many conclusions related this megatrend, such as:

a The formation of new professions and the changes in educational systems necessary for bringing these new professions into effect are the most important of the social phenomena that limit the speed of development of the information society, the knowledge-based economy and civilisation.

b New technologies always give opportunity to get rich – for those, who know how to use this opportunity. However, at the beginning of the era of information civilisation, this phenomenon is particularly significant: the megatrend of changing professions causes another trend, that of *increasing social stratification,* sometimes called *a new divide.*

c The condition of the success of the information society is the *human innovativeness in inventing new professions,* that will provide employment for the majority, not only a small part of the population in the knowledge economy and will thus limit the impact of the new divide.

d Many decades of increased demand for education can be predicted. This concerns education at all levels, but in particular university education and life-long continuing education. Together with other technological trends discussed above, this results in a *trend of multimedial life-long and continuing education.*

The **MEGATREND OF INTELLECTUAL CHALLENGES** might be even more important than the previous two: it is formed by the great challenges concerning the way of understanding the world, resulting from the spread of the information civilisation. The *mechanical way of understanding the world* – as a great machine, turning its wheels with the inevitability of celestial matter – *will be replaced by a new way, systemic and chaotic.* The world will be seen as a great but complex dynamic system, in which there are some laws, but chaotic behaviour, resulting from nonlinear dynamics with strong feedback is likely. A chaotic system is more similar to an avalanche or a tornado than to a big machine; anything can happen there, and small changes in initial conditions can essentially change its path. Everything which used to seem a natural and common sense phenomenon in the old way of understanding the world might be questioned in information civilisation. Thus, we might have many new and great conceptual, intellectual challenges.

This concerns quite basic concepts and problems: the way of understanding markets, economy, democracy, human rights, ethical problems, etc. We shall not discuss this megatrend in any more detail here. However, it should be noted that the subject of this paper – the issue of modelling for knowledge exchange – is strongly related to this third megatrend, though it also results clearly from the definition of the knowledge-based economy and is related to both megatrends discussed earlier. It should be also noted that it is the third megatrend that constitutes the greatest challenge to science policy: without seeing the new knowledge-based economy in new terms, we might miss the most important policy aspects.

2 Diverse concepts of information and knowledge

When discussing the future of knowledge-based economy, in order to respond to its intellectual challenges, we must better understand the meaning of the term *knowledge.* An encyclopaedic definition of knowledge usually stresses at least two aspects of this concept. One understanding of knowledge corresponds to the entire contents of an individual mind, resulting from experience and learning. Another understanding of knowledge stresses its objective and socially utilitarian character: knowledge denotes all information that results from confrontation with the real world, be it incorporated in theoretical reflection or be it not based on a theory, but just useful in applications. In this paper, we are not concerned with the individual, but with the social aspect of knowledge. However, both above definitions are very general, while in the time of the information society and knowledge-based economy we need a more technical definition of knowledge. Recall that the known technical definition of the quantity of information[1] contributed essentially to the understanding of this concept, even if various qualitative aspects of information, such as its security, quality, etc., are of more importance today. However, the concept of knowledge is more demanding that that of information, and we shall consider it from various perspectives: so-called *soft sciences* (humanities and social sciences) versus the so-called *hard sciences* (natural sciences and technology).

Both in soft and hard sciences, we often use the concept of a *model.* This concept is quite general. Besides its traditional social or professional interpretations, it might mean a small representation or a copy of some object, a pattern on which we base a new version of a system, etc. However, in professional, scientific language, the concept of a model denotes *a synthetic description of a part of reality, constructed for a given purpose.*

Following the *falsification* approach in the philosophy of science – see Popper (1983) – we can interpret any *scientific law* as a *model of reality,* valid only subject to falsification attempts (which failed to falsify this model). The Popperian approach to the philosophy of science is today considered strict and demanding, at the same time there are also even more relativistic approaches. For example, Feyerabend (1987) denies any fully objective character of knowledge. However, the most convincing seems the approach of evolutionary epistemology (see e.g. Lorentz, 1965, Wutetits, 1984): objective knowledge and its social transfer are useful for the evolution of humans. We can treat the principle of falsification by Popper as an useful tool for checking the objectivity of knowledge, needed in the evolution of human civilisations. We can summarise the above discussions thus: *models express knowledge while no knowledge is absolute.*

2.1 Concepts of knowledge in humanities and social sciences

Humanities, with their distinction of nomothetic and ideographic sciences, have a quite different concept of knowledge than the so-called hard sciences. There is actually a duality, an opposition of two different concepts of knowledge in humanities. One of them is still a model, although in a rather general, verbal form: *ideal type* (M. Weber) or a *structure* (of G. Levi-Strauss). The opposite concept stresses not model formation, but *hermeneutic understanding* (with a long history of the development of the concepts of hermeneutics, Husserl, Gadamer, Heidegger). This is a very deep counter-position, a dichotomy of basic concepts: *knowledge as models versus knowledge as understanding.* With the emergence of knowledge science, on the verge of knowledge-based economy, we also need a deeper though rational explanation of this dichotomy.

The explanation of this may be obtained from the rational theory of intuition proposed by one of the authors (Wierzbicki, 1997). Without presenting this theory in full, we shall include here a short discussion of the role of subconscious thought, intuition and non-verbal processing by human mind.

We know today that processing images requires circa 10^4 times more processing capacity than processing words[2]. Humanswere already quite capable of dealing with their environment, before developing language, hence had quite well-developedcapacity to process images and other signals in a holistic way; this was related to a subconscious use of multivalued, fuzzy logic, not to binary logic. The development of language started the accelerated evolution of human civilisation, the intergeneration transmission of knowledge. Binary logic was discovered together with language, in order to convince others about necessary actions (by arguments such as: *we must take this action, otherwise very bad things would happen*), but is not naturally related to the description of a real world. This development pushed the original, holistic processing of visual and other information further into the subconscious parts of our mind; we still have the capacity of such reasoning, but we call it *intuition.* From these premises, a rational theory of intuition can be developed, together with practical falsification tests of theoretical conclusions, see Wierzbicki (1997, 2000).

Thus, we need to extend the dichotomy between humanistic models (ideal types, structures) and humanistic understanding (hermeneutics, which can be viewed as intuition formation) to other forms of knowledge and other sciences. We need both the evaluation of various forms of models expressing knowledge and, in order to show the limits of knowledge, a rational theory of intuitive reasoning, if we aim at the development of *knowledge science.* Such a new discipline should not only help us to understand various forms of and limits to knowledge, but also result in a better integration of and transfer of results between diverse disciplines of knowledge and science.

The rational perspective of intuitive reasoning also helps to better understand another distinction between various forms of models expressing knowledge. Hard (nomohtetic) sciences engineering in particular, often use *mathematical models*. However, disciplines such as philosophy, humanities and social sciences – generally called soft (ideographic) sciences[3] – attempt to help in the understanding of an increasingly complex world in various ways. One way is by trying to reach a hermeneutic, intuitive understanding; another way is by using various general types of models – ideal types, structures, other verbal models of difficult issues.

Note that even a verbal discussion of such issues, organised in a structured way, is in fact a model that tries to provide a certain perspective and to increase understanding. The megatrends of the information society, discussed earlier in this paper, provide an example of such a model. We might call such verbal models *humanistic models,* as opposed to mathematical models.

Soft sciences have in the past attempted – sometimes redundantly – to increase the scientific validity of their statements by trying to incorporate mathematical models from hard sciences. Humanistic models however, rely in fact on hermeneutic understanding, at a very high degree of synthesis. From the rational theory of intuition it follows that such degree of synthesis is usually attained by intuitive, holistic perception. Choosing words to formulate such models is difficult, because all words have multiple meanings; on the other hand, it is doubtful whether any attempt to make these models 'harder' by trying to express them in a mathematical form is constructive or useful.

An example of such difficulties is the issue of interpreting the statistical models of data preferred today by social sciences. These sciences, such as economics, business management, quality management etc., tend to be fascinated by the possibilities of computer data processing and by deriving statistical models from these data – with no deeper knowledge of the methodology of forecasting which stresses the dangers of interpreting such models. A statistical model is not a causal model, to specify causes and effects. In such a model we need deeper, fundamental knowledge, external to the statistical data. The best illustration of this is the strong statistical relation between magnetic storms on Earth and solar spots. Based only on statistical data, one could advance a theory that magnetic storms perturb our vision and thus we see spots on the Sun. Such a theory is false, but consistent with the data. In order to correctly specify that solar spots cause magnetic storms, we need the theory of elementary particles emitted by solar eruptions and their impact on the magnetic field of Earth – a deeper theory, external to the data.

This example also illustrates the conclusion that it is important to develop *knowledge science* but it must include *methodology of model formation and interpretation.*

2.2 Concepts of knowledge in hard sciences

Hard sciences have their own forms of knowledge: though all use mathematical models, they are by no means united about the definition of knowledge. We start thus from a specific discipline in information sciences, called *knowledge engineering.* From the perspective of knowledge engineering, knowledge is defined as a *pattern that can be discerned in data.* While this definition is very useful for *data mining* and *knowledge discovery* in large data sets, it is possibly too narrow for broader applications in the knowledge economy. Although knowledge engineering specialists tend to include *models* as a specific forms of *patterns,* the interpretation of this relation in other specialities – including computerised decision support – is just the opposite: *patterns* are considered a specific form of *models.*

We shall follow the second interpretation, since the concept of a *model* is quite general. Besides its traditional social or professional interpretations, it can mean a small representation or a copy of some object, a pattern on which we base a new version of a system, etc. We have already mentioned that in professional scientific language, the concept of a model denotes *a synthetic description of a part of reality, constructed for a given purpose.* The purposes of constructing a model can be diverse, but very often – e.g. in model-based decision support – relate to expressing knowledge about a given situation. The forms of models can be also diverse, starting with a general, verbal form and proceeding to more specific forms of mathematical models.

In a sense, mathematical models today represent the most basic form of expressing and exchanging knowledge in hard sciences, technological knowledge in particular. This is because mathematical models can be easily computerised. We shall discuss this and related concepts in more detail.

3 The importance and typical forms of mathematical models expressing knowledge

A critical element of many scientific investigations – also including such activities as model based decision support – is a mathematical model representing data and relations that are too complex to be adequately analysed based solely on experience and/or intuition of a researcher or a decision maker. Models, when properly developed and maintained, can represent not only a part of the knowledge of a researcher or a decision maker, but also integrate other available relevant knowledge from various disciplines and sources. Moreover, models, if properly analysed, can help their users to extend their knowledge and intuition. Therefore, the quality of the whole cycle of preparing, maintaining and analysing models also determines to a large extent the quality of research conclusions or of a practical decision-making process for any complex decision problem.

There are new developments in modelling methodology and tools that are worth discussing in the context of knowledge integration. In addition to the continuously growing opportunities resulting from the progress in database management and in the foundations of modelling, there are many new opportunities. They have emerged for example, from recent developments in methodologies for and experiences from, model-based decision support systems – see e.g. Wierzbicki et al. (2000) – or from the network-based platform-independent software technologies. New opportunities are also offered by new developments in the applications of World Wide Web for management, policy makers, research and education. Not all such opportunities are yet being efficiently exploited.

There is a need to use all these opportunities to improve the low productivity of model-based work by unifying various representations of a model, facilitating the standardised interfaces to diverse solvers that support different paradigms of model analysis. It is also necessary to improve the knowledge sharing possibilities represented by various models that can be analysed on heterogeneous hardware available in distant locations.

To look at improving such possibilities, we shall first discuss typical forms of mathematical models. They can be divided into *binary models* (based mostly on binary logical relations, e.g. such as patterns discovered in data) and *analytical models,* although this distinction is not always sharp (each model includes logical relations and might include binary ones). However, this general distinction is related in a sense to the basic megatrend of technical integration in information civilisation. This basic trend is based on digitalisation of all forms of information; thus, it is not strange to think that the forms of models expressing computerised knowledge will also be closely related to the most basic, binary form of presenting information.

3.1 The importance and limitations of binary models

There are many examples of this trend to replace all other types of information processing by purely binary processing. Genetic algorithms of optimisation represent one such example; another is the powerful trend towards finding patterns of knowledge in very large data sets. Thus, there is no doubt that *binary forms of computerised models encoding knowledge,* in a sense natural for implementation on digital computers, will become increasingly important in the future. For a computer scientist, who tends to see the world as a giant computer, these forms are the most natural models.

This does not mean, however, that this form of model will become a universal or even a dominant standard, because for all their uniformity, binary models have some essential drawbacks. These include:

- Binary models are in a sense too sharp for representing the real world; multivalued, fuzzy logic expressions and so-called *soft computations* are much more adequate, but already this means a departure from binary models towards analytical models. will not be For more detail on the developments and applications of fuzzy logic – see e.g. Zadeh (1978), Zimmermann (1987).
- Each discipline of science uses its own characteristic collection of analytical models which serve as a kind of global language for specialists in this discipline; it would be disadvantageous to the development of science to replace such 'languages' by a universal, binary model form. We should recall here several examples. One of the disciplines that has contributed most significantly to the understanding and classification of analytical models is *control science.* It described ways of controlling diverse dynamic processeswhich lead to the development of most of modern technology such as automated manufacturing, robotics etc. But con-

trol science also relied on earlier developments of various concepts and models developed in other disciplines, such as the concept of *feedback* coming originally from telecommunications, or models describing dynamic processes coming originally from mechanics. Control science included binary models in the form of the *theory of finite automata,* but is much broader and richer than this theory.

- The processing of binary models requires usually times that grow faster than polynomially (e.g. exponentially) with the dimensions of these models. This is also true of many analytical models. Classical computer scientists usually argue that the processing power of modern computers is growing fast enough not to be troubled by the non-polynomial computation complexity. However, this argument does not take into account the fact that true specialists in complex computations can always find ways to saturate even the most powerful computers. In fact, the advantage of analytical over binary models is the possibility of finding special algorithms – admittedly, usually only approximate and only for limited and specific classes of analytical models – that can process these models much faster than in the universal case (that is always doomed to non-polynomial complexity).
- Finally, we should note that human mind works quite differently than do modern computers and binary models are a very poor approximation of its operation. A better approximation of the human mind are the artificial neural networks, but they are not yet adequate; we can only say that the processing of information in the human mind is certainly parallel and distributed, and certainly more complicated than contemporary artificial neural networks. This argument is also related to the rational definition of intuition discussed earlier in this paper.

3.2 The role and challenges of analytical mathematical modelling

Analytical *mathematical modelling* started as a generalisation of modelling techniques from several disciplines. Models of *operations research* were augmented with a broader methodological reflection of *control science* as well as other disciplines, especially *systems analysis.* Today, the classification of model types in mathematical modelling are well known: *continuous versus discrete, linear versus nonlinear, deterministic versus stochastic, static versus dynamic, open loop versus feedback, single-objective or scalar optimisation versus multi-objective or vector one,* as well as diverse techniques of dealing with all these model types. This powerful body of knowledge can be developed further as one of basic elements of knowledge science, provided we are able to respond to several challenges – all related to the development of the information civilisation and knowledge-based economy. Especially important are the two facets of the megatrend of integration: we need more integration of models used in diverse disciplines and we need clearer principles of using commercially knowledge encoded in models. We shall discuss the first facet here and the second in a further section.

Several challenges can be listed as being related to the facet of integration. The first is the need to integrate, in particular, the results from two disciplines almost separate until now: *knowledge engineering* and *mathematical modelling.* This is related to the former discussion of the binary and analytical forms of models.

The second is to integrate various methodological approaches to the analysis of diverse types of mathematical models used in various scientific disciplines. We recall that mathematical modelling typically uses *model simulation, scenario and sensitivity analysis, single-objective optimisation,* and sometimes *multi-objective optimisation and analysis.* These approaches might be complemented, however, by *multi-objective inverse simulation and scenario analysis,* see e.g. Wierzbicki et al. (2000). However, all these methodological approaches are not widely known by specialists in the various scientific disciplines. While we will discuss these issues in more detail in the next subsection, the related challenge should be noted here : to make the approaches of mathematical modelling and systems analysis known and usable for diverse scientific disciplines.

The third challenge relates to the fact that computerised mathematical models have diverse standards not only between the various scientific disciplines, but also within each discipline. This makes knowledge exchange cumbersome and difficult and it also impedes the comparison of models devoted to the same subject.[4] There initiatives in existence aimed at responding to this challenge, including the idea of *structured modelling language,* see Geoffrion (1989) and further discussions of modelling tools in Wierzbicki et al. (2000). However, it is not only standards of modelling languages that should be further developed, but also the methods and tools related to model analysis should be integrated into such languages. Thus, much more work is yet to be done before this challenge can adequatelymet.

Other challenges relate to that facet of commercialisation of knowledge which requires a better understanding and also the development of clear guidelines for utilising model-encoded knowledge in several domains. However, as it is not limited to analytical mathematical modelling we will discuss it in a further section together with science policy issues.

3.3 An example: computerised decision support systems

Some necessary developments can be best described using the example of model-based computerised decision support systems. Any decision maker, before making final selection of a decision, wants to understand in the consequences of his/her possible decisions a best possible way. Particularly in more complex situations, decision makers typically need help in finding decisions that best correspond to their preferences. These preferences cannot be precisely defined in advance, because they often change while a decision maker is learning about the decision

If science remains segmented into disciplines, we not be able to deal effectively with the multifaceted

problem. Experiences show that such learning is an important element in the development and use of a decision support system for any complex problem, such that the intuition and expertise of the decision maker are not enough for predicting the consequences of various decisions. However, decision makers typically also want to examine the consequences of the decisions that they define, often by using their own intuition and/or experience to modify decisions obtained from analysis or suggested by somebody else.

A modern decision maker is confronted with more complex decision problems than were previous generations, but also has a much better knowledge of decision making processes and access to analytical tools and teams of experts and advisors. Therefore, such a decision maker is not keen to accept the classical ways of decision support which relied on using, as a basis for a decision, a given solution of a mathematical model that is assumed to represent a well structured problem. A modern decision maker needs a decision support system that can be used for various types of analysis, and that can help to extend the decision maker's knowledge about the problem on the one hand, and allow them to use their experience and intuition on the other hand. In order to achieve this, a good decision support system should be composed of two mutually linked parts of a quite different nature:

- **A MATHEMATICAL MODEL** that represents the part of a decision problem for which logical and physical relations exist and have to be handled in model form rather then by the intuition and experience of the decision maker. We shall call this model a core model or substantive model.
- **TOOLS FOR A COMPREHENSIVE ANALYSIS** of the core model. Up to now we have concentrated more on the diverse types of models; here we shall add some comments on the *tools for model analysis.*

Mathematical and computer models are widely used in many areas of science and industry for predicting the behaviour of a system under particular circumstances, when it is undesirable or impossible to experiment with the system itself. The understanding of the system gained through a comprehensive examination of its model can greatly help in finding *decisions* (in managerial terms; in technical terms they might be called *controls*) whose implementations will result is a desired behaviour of the system. The systems that are modelled have diverse characteristics – including the nature of the underlying physical and economic processes, their complexity, size, types of relations between variables etc. There is also a great diversity in the use of models that depends on various factors – such as the nature of the decision making process, the background and experience of users of the models, etc. However, the process of formulating and analysing a model, abbreviated to *the modelling process*, have many similarities even if the systems being modelled are quite different. This type of modelling process is typically composed of problem formulation, model specification, implementation, verification and validation, analysis and management. The modelling process is a combination of craft, art and knowledge, and its quality is critical to any results achieved.

will

problems.

There are two classical approaches to model analysis: *descriptive* (also called *predictive*) and *prescriptive* (or *normative*). Descriptive modelling is used for predicting the behaviour of the modelled system without attempting to influence it. Prescriptive modelling is aimed at providing information about decisions (or controls) that can result in a desired behaviour of the modelled system. In other words, descriptive modelling helps to answer the question '*what will happen if ...*', whereas prescriptive modelling provides answers to the question '*what decisions are best for our purposes'.*

For the prescriptive model analysis, typical tools are *optimisation techniques.* They consist of selecting that decision or solution (in the set of admissible decisions or solutions) that is considered the best in a specified sense. Usually, this sense denotes a solution that results in the best (minimal or maximal) value of a performance index – goal function, objective function, utility or value function, etc. The classical approach assumed that the mathematical model, including the performance index, describes the reality well, including the preferences of the decision maker; thus, it would suffice to optimise the performance index in the model and the best decision would result. Modern approaches question such assumptions at least partially: even if a model describes the reality reasonably well, it usually describes the human preferences rather inadequately; hence the advice not to model preferences and to limit the core or substantive model to knowledge that is as objective as possible. Even so, optimisation might be used for model analysis, but not as the goal of the exercise, only as a technical tool; for example, by using optimisation techniques we might check whether a set of desired outcome values in the model is attainable by admissible decisions, or not.

Descriptive methods of model analysis are useful when *verifying* a model. Not only should a good model represent possibly most objective knowledge while conforming to formal specifications, but also all possible discrepancies between the results of its analysis and the intuitive judgement of model user (a decision maker, an analyst, etc.) should be resolved. Such inconsistencies show that either the intuitive judgement, or a part of model specification (assumptions, data, parameters) is wrong and must be modified. Without such verification, the user will not trust the model, hence the model cannot be used for decision support.

Typical tools for descriptive model analysis are *simulation techniques.* Decision (or controls) are then defined either correspondingly to typical data, or varied randomly, or specified by intuitive judgement of model user; various outcomes defined by the model are computed and presented to the user as results of the simulation. In fact, more advanced simulation techniques mix simulation with optimisation; for example, so-called *inverse simulation* consists of specifying desired outcomes and using optimisation in order to find corresponding decisions or controls.

The usage of simulation and optimisation could be ideally compared as follows:

- In simulation, mode decision variables are inputs and goals are outcomes. This technique is good for exploring the intuition of a decision maker, not only for model verification but also for examining the consequences of applying certain decisions in terms of goals and constraints. Simulation can be considered as an option-focused method of analysis that aims at examining options.
- Optimisation can be considered as a value-focused (or goal-oriented) approach that aims at creating options. Optimisation is driven by the hope of reaching a set of goals (while the basic goal is to optimise the value of a given performance index). Therefore, goals are a driving force and the values of decision variables are outcomes. This is very appealing, but has certain disadvantages. First, not all real objectives are usually included into the goals formulated for the optimisation. Second, an optimal solution is selected between members of the set of admissible solutions. If a decision maker prefers to change this set, a new optimisation problem must be considered.

Therefore, an interchangeable use of both simulation and optimisation techniques has obvious advantages, but a joint implementation of both techniques in classical models is difficult. This has resulted in development of *multiobjective model analysis* methods and *vector optimisation* tools that combine advantages of these two classical approaches.

Advances in methods and tools for specification and analysis of analytical models have allowed for implementations of such models in various policy making problems for which traditional crisp optimisation and/or simulation modelling tools do not offer enough support. Modern model based decision support explores a cluster of enhanced traditional methods combined with multicriteria model analysis. This provides advantages in examining trade-offs between various conflicting criteria and helps to identify attainable goals and decisions which lead to achieving such goals. Various techniques can be used for multicriteria model analysis; best developed are *reference point approaches* that can be considered as a generalisation and enhancement of known *goal programming* techniques. Reference point approaches (see Wierzbicki et al., 2000) include techniques for *soft and inverse simulation* as well as examining *soft constraints*. Decision support systems based on reference point approaches are usually built from a modular collection of software tools, to which belong:

Model generation tools. Generation of a core model or a substantive model (which is a representation of all logical and physical relations between variables representing the decision problem being examined, without including models of the preferential structure of a decision maker) requires special tools. Various *modelling languages, model generators,* etc., can be used for this purpose.

Model analysis tools. As discussed above, model analysis can have various tasks and tools: simulation, optimisation, vector optimisation, soft and inverse simulation, examination of soft constraints, etc.

COMPUTATION TOOLS. Model analysis often requires solving a series of auxiliary optimisation or simulation problems; this in turn requires robust and efficient solvers that can handle the related computational tasks in a way that is transparent to the user.

The diversity of such modelling tools and types of model raises the issue of *model standardisation.* Many research institutions develop analytical models that are of broader interest. However, there are no standards for a model specification and analysis. Therefore, not only is much work required for the development of models but also the analysis of a model is often restricted to a limited number of approaches for which tools appropriate for the particular model class are available.

To understand why a standardisation is a possibility, we have to recall the evolution of database management theory and technology. The data management revolution occurred in response to severe problems with data reusability associated with file-processing approaches to application development. The need to share data resources resulted in the development of data base management systems that separate data from the applications that use data. Following this historical example, model management systems can be developed. Models could not only be developed in a much more efficient way but they could also be used in a more efficient way, if standards for model specification and analysis were agreed upon. We observe thus that model management is at about the same stage of evolution as data management was during its transition from file processing to database processing. It should also be noted that data management has recently been subject to a fundamental change, related to the data warehouse concepts. The ability to capture data from multiple operational source databases, to date data consistently, to retrieve it efficiently across many different dimensions has always been a key issue for timely delivery of useful information needed for actual decision making; data warehouses respond to such needs.

The comparison of data bases and analytical models clearly illustrates the challenge: what moved database technology forward – voluntary de facto standardisation around a rigorous, principled representation formalism of great generality – is what we need to move model management forward. Standardised interfaces for model analysis would in turn, allow much broader access of both institutions and individuals to a vast amount of information and knowledge that is currently available mainly for researchers and experts. Non-specialists will not learn detailed ways of analysing diverse model types, but many may be willing to learn user-friendly ways of analysis of various models that can be made publicly available. Then a broad access to knowledge encoded in models can greatly contribute to the education of societies and can help in public discussions of various issues, such as social security reforms, population ageing, climate change, air quality, etc. Finally, standardisation of models is necessary for integration of models from diverse disciplines, thus for advancement of trans-disciplinary and interdisciplinary knowledge.

4 Science policy issues

All of the above discussions have some implications for science policy. We list only some of these.

1 Essential for a good formulation of science policy on the verge of knowledge-based economy, particularly in terms of the megatrend of intellectual challenges, is a deeper understanding of the concept of knowledge, in several aspects such as:

- We need further analysis and comparison of diverse concepts and definitions of knowledge, their interdisciplinary interdependence, to formulate science policy for stimulating interdisciplinary research that will become increasingly important in knowledge-based economy.
- In particular, we need stimulation of development of new economic models of production and new economic theory of knowledge-based economy.

2 Essential for issue for science policy in knowledge-based economy is a good balance of knowledge utilisation in various domains:

- *Private domain* concerns issues of intellectual property rights for model-encoded knowledge as well as the development of new protection mechanisms (since it is difficult to imagine patenting model-encoded knowledge).
- *Commercial domain* concerns issues of marketing and commerce in model-encoded knowledge.
- It should be stressed, however, that for all the development of commercialisation of knowledge, much of it would certainly remain in *public domain*. Free access to at least some parts of knowledge is advantageous for the evolution of human civilisation and many scientists will insist on making public some of their results. But the issues of knowledge in public domain are by no means simple. The distinction between the private, commercial and public domains is a deeply ethical question. The discussion of this question will remain one of main intellectual challenges related to the third megatrend of intellectual challenges.

3 Societies and countries who support a faster development of knowledge-based economy will have a better position in the information civilisation. Thus, science policy should specifically support those areas of research that might contribute most significantly to the development of knowledge-based economy. We shall mention here some of such areas:

- *Knowledge science* – as a transdisciplinary synthesis of diverse approaches to knowledge formation, based not only on knowledge engineering, but including also on other forms of model-encoded knowledge such as in mathematical modelling, as well as basic philosophic reflection on the concepts of knowledge. An essential issue in knowledge science is the dichotomy between *knowledge as models versus knowledge as hermeneutic or intuitive understanding;* this issue should be studied further in order to deepen the foundations of knowledge science.

- Since knowledge is becoming increasingly more commercialised, another essential issue in science policy is the support for the development of *standards of knowledge encoding,* including modelling languages integrated with methods and tools related to the analysis of models of diverse types encountered in various scientific disciplines.

5 Conclusions

There are many conclusions that might be drawn from this quite general discussion of the concepts of megatrends of information civilisation and the role of modelling for knowledge exchange; only some of them are stressed here:

- A new understanding of the coming civilisation era is extremely important for science policy; such an understanding might be helped by the megatrends of information civilisation, discussed above as a simple model of this coming era.
- Science policy should include a better balance of knowledge utilisation in various domains: private, commercial and public.
- Science policy should also support the development of betters tools for knowledge exchange, including such as standards of knowledge encoding and modelling languages.

References

(1965) K. Lorentz: *Evolution and Modification of Behavior: A Critical Examination of the Concepts of the 'Learned' and the 'Innate' Elements of Behavior,* The University of Chicago Press, Chicago-London 1965

(1978) L. Zadeh: Fuzzy sets as a basis of a theory of possibility. *Fuzzy Sets and System*s 1, 3-28, 1978

(1980) A. Toffler: *The Third Wave.* W.Morrow, New York 1980.

(1983) K. Popper: *Realism and the Aim of Science.* Hutchinson, London 1983

(1983) A.P. Wierzbicki: Changing Cultural Paradigms. *Options,* Nr 4, 1983, IIASA, Laxenburg, Austria.

(1984) F. M. Wuketits: Evolutionary epistemology – a challenge to science and philosophy. In F. M. Wuketits, ed.: *Concepts and Approaches in Evolutionary Epistemology.* D. Reidel Publishing Co., Dordrecht 1984

(1987) P. Feyerabend: *Farewell to Reason.* Verso Publishing Co., New York 1987

(1987) H.J. Zimmermann: *Fuzzy Sets, Decision Making, and Expert Systems.* Kluwer Academic Publishers, Dordrecht, 1987.

(1988) A.P. Wierzbicki: Education for a New Cultural Era of Informed Reason. W: J.G. Richardson (ed.): *Windows on Creativity and Invention,* Lomond, Mt. Airy, Maryland 1988.

(1989) A. Geoffrion: The formal aspects of structured modeling, *Operations Research,* **37** (1), 30-51, 1989.

(1997) A.P. Wierzbicki, K.B. Wydro, A. Zieliñski (in Polish): The era of information civilization and the developmental role of telecommunications, *Przegl1d Telekomunikacyjny i Wiadomoœci Telekomunikacyjne* No. 5, 1997, Warsaw, Poland.

(1997) A.P. Wierzbicki: On the Role of Intuition in Decision Making and Some Ways of Multicriteria Aid of Intuition. *Multiple Criteria Decision Making* **6**, 65-78.

(1999) A.P. Wierzbicki: *Megatrends of Information Civilisation.* International Conference on Research for Information Society, Warsaw 1999.

(2000) A.P. Wierzbicki, M. Makowski, J. Wessels: *Model-Based Decision Support Methodology with Environmental Applications.* Kluwer- IIASA, Dordrecht-Laxenburg 2000.

(2000) A.P. Wierzbicki: *A Rational Theory of Intuition in Decision Making and its Relation to Negotiation Techniques.* Conference on Group Decisions and Negotiation, Glasgow 2000.

Notes

1 Due to Shannon: the amount of information is equal to the binary length of its code.

2 Transmitting images requires 102 times more transmitting capacity than transmitting words; processing times grow usually at least with the square of the amount of data.

3 In a positive sense, as illustrated by the following principle that was formulated by H. Barnett (in a discussion at IIASA, 1980): *Hard models – soft thinking, soft models – hard thinking.*

4 An example relates to models of climate, particularly global ones: due to the lack of compatibility of modelling standards, it is extremely difficult to compare the assumptions and conclusions of various studies of global climate.

Trans-Disciplinary Knowledge Exchange

Etsujiro Shimemura
President, Japan Advanced Institute of Science and Technology (JAIST)

Yoshiteru Nakamori
School of Knowledge Science, JAIST

Abstract

Human society is becoming increasingly complex. If science remains segmented into specialised disciplines, we will not be able to deal effectively with the multifaceted problems, which we are now facing. We need a new integrative science that is founded on the deep understanding of humanity and society. To this end, School of Knowledge Science was established in 1998 at the Japan Advanced Institute of Science and Technology to lead the coming knowledge society.

In view of this initiative which aims to discover both theoretical and practical principles of knowledge need, the school has embarked upon new knowledge management (i.e., the management of creating new knowledge and integrating it with existing knowledge), thereby developing new knowledge systems for decision making and problem solving.

The school has enlisted not only natural scientists and engineers but also social scientists and humanities scholars. These faculty members conduct research into innovative methods for solving complex problems, and man-computer systems that support such problem-solving activities.

The school also offers master's and doctoral programmes to educate both professionals (e.g., project-team leaders and knowledge engineers) and knowledge scientists and to equip them with knowledge-creating methods such as fieldwork, statistical analysis, simulation, knowledge engineering, etc. They are expected to become the pioneers of the knowledge society.

In this presentation, we will introduce an ambitious programme for establishing knowledge science, including current achievements, present conditions, difficulties and future problems. Discussion will be extended to include the possibility of trans-disciplinary knowledge exchange utilising information and communication technologies.

Contents

Skills in High Demand: Bioinformatics

Sylvia J. Spengler, Ph.D.
Centre for Bioinformatics and Computational Genomics
Lawrence Berkeley National Laboratory

Abstract

Bioinformatics is a field that draws on computer science, engineering, statistics and biology. What distinguishes bioinformatics from more traditional computational biology is its emphasis on large data sets of high complexity. This necessitates tailor-made training programs in bioinformatics. It is not only interdisciplinary training that is necessary to develop people with skills in bioinformatics, but also institutional changes. This paper draws lessons from the experiences in training bioinformatics in the US.

I take bioinformatics to mean the broad range of computational sciences that are brought to bear on biological problems. What distinguishes bioinformatics from more traditional computational biology is its emphasis on large data sets of high complexity. If folding a protein is the classic question of computational biology, then bioinformatics seeks to develop the tools for predicting the structure of all gene products.

Bioinformatics is a field that draws on computer science, engineering, statistics and biology. It is necessary for those trained in the more traditional computational sciences to acknowledge one basic reality: biology is hard because biology is different. Each living organism is unique, characterised by a history, the impact of chance, and a high digital information content. The uniqueness of organisms, even in bacterial colonies that might seem to be all identical at first glance, means that there really is no law of large numbers. Nevertheless, modern biochemistry and molecular biology have managed to accumulate an extraordinary amount of data at many levels of resolution about living systems.

In 1998, the US National Science Foundation funded a programme in mathematics and Molecular Biology (PMMB) at the University of California, Berkeley. This was a Centre without Walls, a multi-university group of investigators who were interested in doing research at the interface between computation and, specifically, molecular biology. Chemists, mathematicians, computer scientists, statisticians, and biologists were among

We need tailor-made training programs in bioinformatics.

the founding members. It was clear from the beginning that we needed to develop a training programme that would support students in learning enough of more than one science to be skilled at the interface. We devised a National Fellowship program through PMMB, available for both graduate students and postdoctoral fellows, whose research would be at the interface. The fellowship programme has been cloned by the Department of Energy and the Sloan Foundation, bringing significantly higher stipends and visibility than a relatively local programme could produce. However, it must be recognised that the National Science Foundation took a major leadership position in enhancing training and research at the interface of biology and computational sciences. The effectiveness of this programme has been significant, in that many of the early fellows are now faculty members at institutions around the world. Nevertheless, it must be admitted that most of the positions are in traditional departments, although there are some exceptions.

This raises the reality that it is not only interdisciplinary training that is necessary to develop people with skills in bioinformatics, but also institutional changes. Until very recently, interdepartmental appointments were a rarity, as were Bioinformatics Departments or Programmes. To encourage universities in the US to take a more proactive position and encourage bioinformatics, the Borroughs Wellcome Foundation developed a program on Science at the Interface, held a competition, and established five Borroughs Wellcome Foundation Programs around the country. It was seen as so effective that a second round of applications was opened and funded. Student exchanges, supervisors from both biological and computational science, fellows meetings, national meetings all are supported through these programmes, in addition to interdisciplinary education of varying designs.

All of this occurred before the largest player in biomedical funding in the US, the National Institutes of Health, decided to take a long hard look at the situation. Harold Varmus, the then director of NIH, drew together a stellar, interdisciplinary group to look at the necessity of integrating computational science and support with the rising data flow from genome projects, arrays, combinatorial approaches to therapeutics and an assortment of other high throughput techniques in biomedicine. The findings, summarised in the 1999 Biomedical Information Science and Technology (BISTI) report, called for Centres of Excellence in Biomedical Computing, basic research in biomedical computing, research into information storage, curation, analysis and retrieval, support for a scaleable national computer infrastructure, all wrapped in a full menu of education.

But the questions remains of exactly what elements should be part of the education, how rapidly can institutions develop effective programmes, and not least, how to avoid eating the seed corn of bioinformatics, by having all of the top people drawn into the commercial world from academia. What has seemed to work best so far has been to have individuals education to the Bachelor level in a standard discipline, whether biology or computational science, and then to acquire cross-disciplinary experience at the graduate and postdoctoral level. One new program, focused at the Masters level, has just started at the Keck Graduate Insitute, which admitted its first class this fall (2000). Perhaps the concern over curriculum development is similar to that which must have been part of other cross-disciplinary areas, such as biochemistry or biophysics. However, these disciplines did not face the incredible market that exists for bioinformatics skills and expertise. It is not only at the senior faculty level that the lure of exceptional high salaries exists. Even students with a modicum of computational science and a bit of biology command significant salaries at a variety of new biotechnology companies, as well as bioinformatics companies. One bright spot is that there does appear to be a two-way door between industry and academia in bioinformatics, so that individuals go each way, depending on their personal inclination and situation.

The major problem faced by independent groups, such as the Centre for Bioinformatics and Computational Genomics, beyond holding good people, is the problem of long term funding. Both the National Institutes of Health and the National Science Foundation have established long term funding programs in the area of bioinformatics, but the Department of Energy remains with much shorter funding periods. Universities are in a much better position here, because departments have assured budgets and positions. But what a centre can achieve, fairly rapidly, is a high density of skilled people and interesting problems to tackle.

One of the major drivers for the need for bioinformatics skills has been the decision of US funding agencies to require that data be made publicly available, coupled with the explosion of the internet and the world wide web as the platform for access to data. The size and complexity of these data are far beyond the simple visual images that used to constitute data for a paper. The Department of Energy and the National Institutes of Health established policies for data sharing and materials sharing early in the Human Genome Project. These agreements were the basis for the later international agreements for production sequencing. It really does require funding agency support for such initiatives to hold sway. Journals also play a very important role in mandating the deposition of data in public databases before a paper can be published.

Annotation, the making sense of sequences that actually enhances the value of the string of letters, can expand both the size and complexity of the data by several orders of magnitude. This is where the understanding of the biology needs to go hand in hand with the computational skill. In many instances, this understanding is accomplished by mixing people with different skills in the same laboratory. This 'sitting next to a biologist is a computer scientist' approach will probably remain true until understanding the use of computers becomes part and parcel of being a biologist. Even in this circumstance, there will still be a need for those who understand both sides of the street at a deeper level. That will remain the territory of bioinformatics.